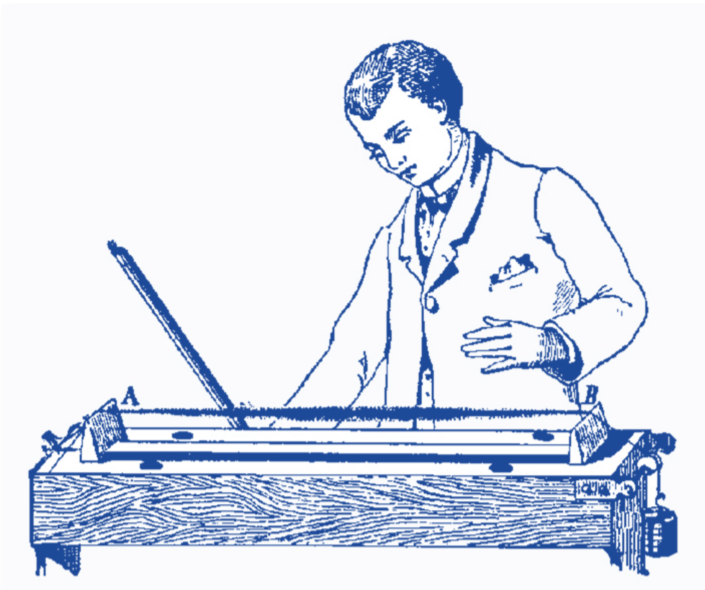
DRAFT

Proceedings of the Stockholm Music Acoustics Conference 2023

SMAC 2023

Presented as a track within the conference
Sound and Music Computing 2023

https://smcnetwork.org/smc2023/



Preface

This volume presents the proceedings of the *Stockholm Music Acoustics Conference 2023* (SMAC), taking place on 14–15 June 2023 in Stockholm, Sweden. SMAC was premiered at KTH in 1983, and has been organized every tenth year since then. This conference is intended for academics, music performers and instructors interested in the field of Music Acoustics. It brings together experts from different disciplines, to exchange and share their recent works on many aspects of Music Acoustics, including instrument acoustics, singing voice acoustics, acoustics-based synthesis models, music performance, and music acoustics in teaching and pedagogy.

This time, our multidisciplinary conference is organized on a smaller scale than earlier, as a track within the 2023 *Sound and Music Computing Conference*, at KMH Royal College of Music and KTH Royal Institute of Technology. Our warm thanks are due to the SMC Network for hosting SMAC in the framework of SMC, and many thanks to all presenters and co-authors for participating. We hope that you will enjoy learning of the new science presented here, and meeting with colleagues and friends, old and new.

This is a draft version of the proceedings, with the papers simply given in the order that they were presented. A citable publication will be produced after the conference.

Sara D'Amario *Anders Friberg* *Sten Ternström*

Track chairs, Editors

Content

Session: Strings 1

1 *Jim Woodhouse*
KEYNOTE: Mapping Playability - The Schelleng Diagram and its Relatives for Bowed Strings and Wind Instruments

9 *Ewa Matusiak* and *Vasileios Chatziioannou*
A Comparison of Friction Models for Bow-String Interaction Based on Experimental Measurements

17 *Masao Yokoyama*, *Yuki Hori*, *Shouta Tani*, *Amane Takei* and *Genki Yagawa*
Sound Analysis of Violin via Neural Network and Numerical Simulation

Session: Strings 2

23 *G O. Paiva* and *Jl Le Carrou*
Numerical Investigation of a Plucking Position Estimation Method in the Time Domain

29 *Ailin Zhang* and *Lei Fu*
Effect of Soundpost Position Adjustments on the Perceptual and Vibrational Qualities of the Violin

37 *Jim Woodhouse*
Bowed String Transients: Enhanced Modeling of Friction and Finite-Width Bows

43 *Jacob McQuillan* and *Maarten van Walstijn*
Helical Spring Mode Grouping

51 *Mauricio Rodriguez*, *Josie Moran*, *Inti Wiarco*, *Patricia Guevara*, *Jorge Vargas* and *Jesús A. Torres*
Estimation of the Elastic Properties of a Violin Top Plate Using Finite Element Models

Session: Wind Instruments

55 *Song Wang*, *Esteban Maestre* and *Gary Scavone*
Characterization of Single-Reed Instrument Sound Generation Based on Ffowcs Williams-Hawkings Analogy

63 *Christopher Clarke*, *Samuel Goh*, *Da Yang Tan*, *Vincent Tan* and *Jer-Ming Chen*
Assessment of Performance Characteristics of the Sheng (Chinese Mouth-Organ)

68 *Hsien Han Ng*, *Da Yang Tan*, *Teck Seng Ng* and *Jer-Ming Chen*
Acoustic Impedance Spectrometry of the Dízi (Chinese Transverse Flute): Fabrication and Fingerings

72 *Alberto Acquilino*, *Ninad Puranik*, *Ichiro Fujinaga* and *Gary Scavone*
Detecting Efficiency in Trumpet Sound Production: Proposed Methodology and Pedagogical Implications

80 *Enis Ukshini* and *Joris Dirckx*
Transversal Bending is an Essential Aspect of Saxophone Reed Dynamics

88 *Peter Pabon*, *Wilco Kamminga* and *Geert Berghs*
Trombone Range Profile and Voice Range Profile

Session: Percussion Instruments

95 *Geert Karman*
The Reflection Wave and Loudness Variations in Harpsichords

103 *Zehao Wang* and *Miller Puckette*
A Fast Algorithm for the Inversion of the Biharmonic in Plate Dynamics Applications

111 *Henna Tahvanainen*
On the Acoustics of the Concert Kantele

118 *Leonardo Fuks*
Is a Woodblock an Aerophone? A Discussion Among Acoustics and Organology

Session: Performance

126 *Andre Almeida*, *Weicong Li*, *Emery Schubert*, *John Smith* and *Joe Wolfe*
MIPCAT - A Music Instrument Performance Capture and Analysis Toolbox

134 *Weicong Li*, *Andre Almeida*, *Emery Schubert*, *John Smith* and *Joe Wolfe*
How Clarinettists Convey Emotion in Music Playing: Measuring Players' Gestures as well as Signal Parameters Aspects Using an Innovative Toolbox

140 *Reuben Walker*, *Mario Fleischer*, *Marie Bieber* and *Dirk Mürbe*
Development of Vocal Vibrato Measures During Conservatory Training - A Longitudinal Retrospective Study

146 *Sara D'Amario*, *Sten Ternström*, *Werner Goebl* and *Laura Bishop*
Impact of Singing Togetherness and Task Complexity on Choristers' Body Motion

Session: Voice 1

151 *Filipa M.B. Lã*
KEYNOTE: Technology-Based Real-Time Visual Feedback in the Education of Singers

152 *Giorgos Dedousis*, *Areti Andreopoulou* and *Anastasia Georgaki*
The Impact of Room Acoustics on Choristers' Performance: From Rehearsal Space to Concert Hall

159 *Hideki Kawahara*, *Ken-Ichi Sakakibara*, *Kohei Yatabe*, *Mitsunori Mizumachi* and *Tatsuya Kitamura*
Singing Voice Range Profiling Toolbox With Real-Time Interaction and its Application to Make Recording Data Reusable

Session: Voice 2

167 *Malte Kob*, *Jithin Babu Pozhamkandath Thilakan* and *Nathalie Henrich Bernardoni*
Vocal Tract Impedance at the Mouth - From 1995 to Today – A Tribute to Joe Wolfe and John Smith

174 *Georgios Papadimitriou*, *Areti Andreopoulou* and *Anastasia Georgaki*
Preliminary Acoustic Analysis of Articulation Differences in Spoken and Sung French Language by Greek Classical Singers

182 *Gerardo Acosta Martinez* and *Helena Daffern*
A Pilot Study of Vocal Vibrato Incorporating Nonlinear Time Series Analysis

190 *Mauro B. Fiuza* and *Filipa M.B. La*
Phonation and Collision Threshold Pressures Associated with Menopause

MAPPING PLAYABILITY: THE SCHELLENG DIAGRAM AND ITS RELATIVES FOR BOWED STRINGS AND WIND INSTRUMENTS

Jim Woodhouse
Cambridge University Engineering Department, Trumpington St, Cambridge CB2 1PZ, UK
`jw12@cam.ac.uk`

ABSTRACT

Schelleng's diagram famously shows the region within which it is possible to sustain steady Helmholtz motion in a bowed string. It is a two-dimensional subspace of the player's parameter space, and gives a graphical overview of an important aspect of "playability". Guettler's diagram considers a different two-dimensional subset, and shows a region within which it may be possible to execute a "perfect transient" in a bowed string. The idea of representing playability via scans of two-dimensional subspaces of control parameters can be extended to other instruments. For reed woodwind instruments, the two main parameters governing steady note behaviour are the blowing pressure and the reed gap, controlled by the player via their bite force. This plane can be populated using simulation models, revealing a wedge-shaped region reminiscent of the Schelleng diagram. Examples will be shown for the clarinet and soprano saxophone. For a brass instrument, the corresponding pair of major parameters are blowing pressure and lip resonance frequency. Again, this plane can be populated via simulations, revealing the conditions (from a given initial transient) leading to the various possible playable notes. Examples will be shown for the trombone. Many other examples of this strategy are possible: it is a powerful way to explore and communicate important aspects of playability.

1. INTRODUCTION

The idea of "playability" can be applied to many, if not all, musical instruments. The player has certain control parameters that they must learn to manipulate, in order to achieve the musical effects they are seeking. Usually, these parameters must be kept within certain ranges to avoid unwanted sounds. Quite often, there are interactions between the parameters so that the player must learn to coordinate the way they change two or more parameters. Finally, the allowed region in the parameter space might vary between different notes on an instrument, or from one instrument to another. Such differences make important contributions to judgements that one instrument, or one note, is "easier to play" than another.

Understanding the physical origins of these playability differences, and perhaps giving guidance to players or instrument makers about how playability could be improved, has been a major goal of research in musical acoustics for many years. The majority of this research work has been directed towards instruments capable of producing a sustained tone: the bowed strings, and the wind instruments. A familiar example of a playability problem is given by the notorious "wolf note" in the cello. The physics behind that particular phenomenon is quite well understood, but there are many others.

The focus of this article is on some aspects of playability in bowed strings and winds which can be investigated via modelling and simulation, and for which important patterns can be encapsulated in diagrammatic form. Further detail on the material can be found in Chapters 9 and 11 of `https://euphonics.org`

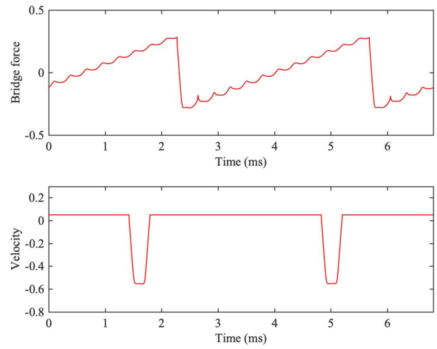


Figure 1. Simulated example of Helmholtz motion of a bowed string. Upper plot: bridge force; lower plot: string velocity at the bowed point.

2. BOWED STRINGS

The thread of research lying behind this article began in the context of bowed strings. Helmholtz was the first to describe the rather unexpected way that a violin string normally vibrates. Figure 1 shows a simulated example of this "Helmholtz motion". The upper plot shows the waveform of force exerted by the vibrating string on the bridge of the violin. The sawtooth shape arises because the motion involves a "corner" that travels back and forth along the string. Whenever this Helmholtz corner reflects at the bridge, it creates a jump of force. The lower plot shows the velocity waveform of the string at the point where the bow interacts with it. Stick-slip vibration results in a pulse-like waveform. Most of the time the string velocity is equal to the bow speed, producing the flat top part of the waveform. Short episodes of slipping result in pulses of negative velocity.

However, the violin doesn't always produce the musical sound a player is after: a bowed string is capable of vibrating in many other ways. In a famous study in the early 20th century, Raman [1] gave an ingenious argument that allowed these other types of bowed-string motion to be described and classified. He argued that they can all be described in terms of travelling "Helmholtz corners". The difference was that Helmholtz motion had a single corner, but the undesirable types of motion had more than one corner.

The simplest example would have two corners, giving string motion with two episodes of slipping per cycle, rather than a single episode. This kind of "double-slipping motion" is sometimes described by violinists as "surface sound". The corresponding string velocity waveform has two pulses in each cycle.

An instrument like the saxophone can exhibit a very similar type of pressure waveform [2]. This is no coincidence: there is a close analogy between a bowed string and a clarinet or saxophone (see for example Chapter 9 of Chaigne and Kergomard [3]), and Raman's ingenious argument can be applied equally well to reed instruments.

2.1 The Schelleng diagram

When a violinist is trying to control the sound of a single, steady note, they have three main parameters to think about: the bow speed, the force with which they press the bow against the string, and the position of the bow's contact point on the string. Forgetting for a moment about the bow speed, there is an interesting interaction between bow force and bow position. In the 1970s, John Schelleng [4] pointed out that if we represent these two parameters pictorially by a point in a plane, then steady Helmholtz motion is only possible if that point lies within a wedge-shaped region of the plane.

For a given bow position, there is a range of allowed bow force. Below a minimum bow force, you get double-slipping motion; above a maximum bow force you get some kind of raucous "crunch". These bow force limits vary with the bow position, leading to the wedge-shaped region shown schematically in Fig. 2. This "Schelleng diagram" encapsulates an important aspect of the "playability" of a violin: a player has to learn to control the combination of bow force and bow position so as to stay within the wedge.

2.2 The Guettler diagram

Knut Guettler was a virtuoso player and teacher of the double bass who was interested in whether theoretical models and computer simulations could tell him things that would be useful in his teaching. Double bass players have a particular problem: some of the notes they play have such low frequencies that they can't afford a bowing transient that takes 10 or 20 period lengths to settle into Helmholtz motion: a short note may be over by then! So Guettler set himself the task of understanding what kind of bow gesture a player needed to perform in order to get a "perfect start" in

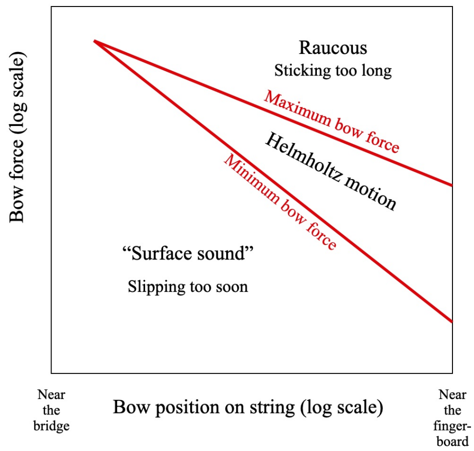


Figure 2. Schematic version of Schelleng's diagram.

which the Helmholtz sawtooth waveform was established right from the first slip of the string over the bow-hair.

This realisation led Guettler to study a particular family of transient gestures. The bow starts in contact with the string, and the force is held constant while the bow is accelerated from rest with a chosen value of acceleration. Guettler then used the simplest available model of bowed string motion to pursue his agenda of finding the conditions under which a "perfect start" was possible from one of these constant-acceleration gestures. He assumed an ideal textbook string, terminated in mechanical resistances. This simple model goes back to Raman, and it is essentially the same model that Schelleng used in his discussion of bow force limits for steady Helmholtz motion.

In a *tour de force* of analysis, Guettler was able to track the behaviour through the first few period-lengths after the first release, and he identified four things that might go wrong with the desired sequence of events [5]. For each of the four, he was able to find a criterion that would decide success or failure. All four criteria take the form of a critical value of the ratio of bow force to bow acceleration. For two of them, the ratio must be bigger than the critical value, while for the other two it must be less than the critical value.

There is a simple way to represent the result, illustrated in Fig. 3. Each criterion corresponds to a straight line in the acceleration–force plane. The slopes of these lines are determined by the four critical values. So there are four radial lines in that plane, and for a perfect start we need to be above two of these lines (shown in blue), and below the other two (shown in red). Unless the criteria are inconsistent, the result is a wedge-shaped region in the plane (shaded yellow here) within which a perfect start might be possible. Such plots are now known, naturally enough, as "Guettler diagrams". This particular example is computed using the frequency and typical impedance of a violin G string (196 Hz and 0.363 Ns/m respectively), and the string is assumed to be undamped. The chosen bow position has $\beta = 0.13$ (the parameter β specifies the position of the bowed point as a fraction of the string length).

All four of Guettler's boundary lines move in a rather

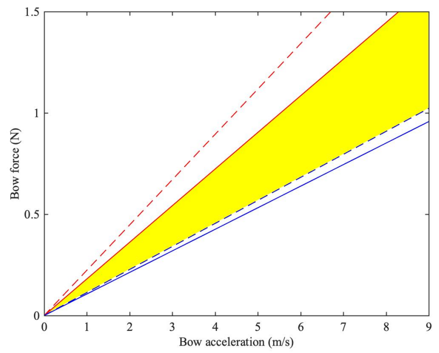


Figure 3. Example of Guettler's criteria for a perfect start. For a given bow acceleration, the force must lie above the two blue lines, and below the two red lines: in other words, it must lie in the shaded wedge-shaped region.

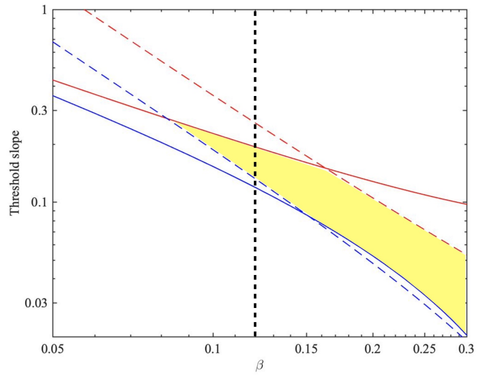


Figure 4. The variation with bow position β of the slopes of the four lines from Fig. 3, using the same line colour convention. In terms of reference [5], the two solid lines represent Guettler's equation (8b), the dashed red line is for his equation (10b) and the blue dashed line is for his equation (12). The region shaded in yellow is where a perfect start might be possible.

complicated way when the bow position β changes. An example of the variation of the slopes of the four lines is plotted in Fig. 4, using the same line colours and types as in Fig. 3. The vertical dashed line marks the value of β used in the Guettler diagram of Fig. 3, and this may help to understand Fig. 4. Moving upwards from the bottom, the first intersection with this black line is the solid blue line, followed by the dashed blue line. These indicate that the solid blue line in Fig. 3 has the lowest slope, followed by the dashed blue line. In a similar way, the next intersection is with the solid red line, and finally with the dashed red line. These are the lines with the highest slopes in Fig. 3.

Logarithmic scales have been used for both axes in Fig. 4, to highlight an intriguing parallel with the Schelleng diagram. The Schelleng diagram shows that for a given bow *speed*, there are limits on the bow force in order for Helmholtz motion to be possible. If β is decreased, both limits increase, and they get closer together and eventually meet. The new diagram says that for a given bow *acceleration*, there are limits on the bow force in order for a perfect start to be possible. These limits, too, increase as β decreases, and get closer together and eventually meet. The pattern is more complicated than the Schelleng diagram, because there are two criteria for each of the upper and lower limits (shown in solid and dashed lines), and the lines cross so that all four play a role in determining the allowed region for some values of β.

Using a computer-controlled bowing machine, Galluzzo [6] made measurements of the Guettler diagram for the open D string of a cello. A typical example of his results is shown in Fig. 5. A 20×20 grid of points in the acceleration–force plane have been tested. Black pixels connote cases that did not lead to Helmholtz motion within an allotted time, while for the others a colour scale is used to indicate the length of transient (in period lengths) before Helmholtz motion was established.

Measurements like this provide an excellent basis for comparison with simulation models. The quest for a fully realistic simulation for a bowed string has been a long one, and is still incomplete – see [7] for a detailed discussion. An example of a simulated Guettler diagram with a "state

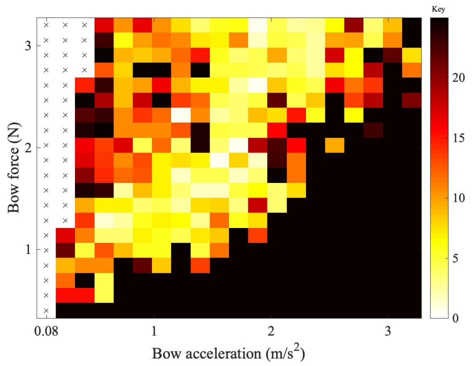


Figure 5. Measured Guettler diagram for the open D string of a cello, colour-shaded to show the length of the transient before Helmholtz motion was established.

of the art" model is shown in Fig. 6. Both the measurement and the simulation show a region of coloured pixels that resembles Guettler's predicted wedge, but the details are rather different. Both cases show "speckled" texture, probably pointing to the fact that a bowed string is close to the mathematician's criterion for chaotic behaviour: repeat measurements of a Guettler diagram scan give slightly different speckle details every time, and this kind of "sensitive dependence" is a ubiquitous phenomenon within chaos theory.

3. REED INSTRUMENTS

The Schelleng diagram and the Guettler diagram both give pictorial information about a two-dimensional subspace of the player's control parameter space. Turning now to wind instruments, it is interesting to ask whether comparable subspace images can be found. Examples will be shown relevant to reed woodwind instruments, and to brass instruments. As with the bowed-string diagrams introduced by Schelleng and Guettler, these examples will be based on

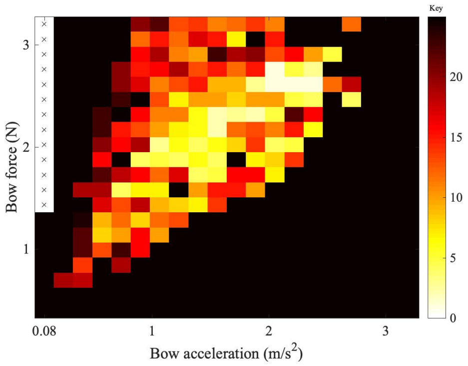


Figure 6. Simulated example of a Guettler diagram using parameter values matching the measurement in Fig. 5, and plotted in the same format.

the simplest theoretical models that capture at least some of the underlying physics. There is no claim that these models correctly represent every detail of real instruments and their players: extending the approach to more realistic models is a challenge for future research.

3.1 The pressure-gap diagram

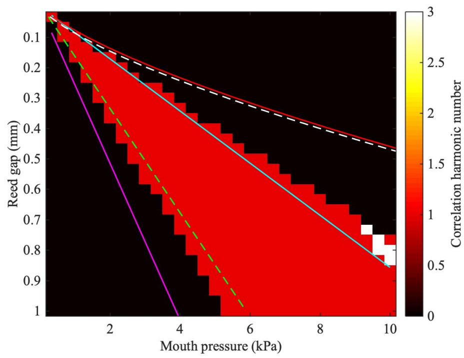
The control parameters available to a reed instrument player are, of course, quite different from those of a violinist. But there is a promising suggestion for a two-dimensional subspace of "playability" parameters in work by Almeida et al. [8]. They plot the results of measurements of clarinet playing, using an artificial mouth, in a plane whose axes are blowing pressure and bite force against the reed.

A slightly modified version of this plane can be used in simulation studies. Blowing pressure is already a parameter in the standard model of the excitation of a reed instrument (see for example Chaigne and Kergomard [3]). Bite force as such does not appear in these models, but probably the main effect of the bite force is to reduce the size of the gap between the reed and the rigid part of the mouthpiece. No doubt the reed resonance frequency and damping are also affected, but for a first attempt at a playability diagram for reed instrument, the pressure–gap plane is a promising candidate. Since the gap varies inversely with the bite force, the gap axis will be plotted in the reverse direction so that bite force increases upwards in the diagrams for consistency with the plots of Almeida et al. [8].

3.2 Results for the clarinet

A clarinet simulation model will be used to provide examples. The instrument tube is modelled using a modal fit to the published input impedance of a clarinet fingered for its lowest note, with all holes closed, taken (by permission) from the website `https://newt.phys.unsw.edu.au/music/clarinet/`. Parameters of the mouthpiece model are taken from Dalmont and Frappé [9]. The reed resonance frequency is 2.8 kHz and its Q-factor is 2.5.

This model has been used to run various 30×30 grids in the pressure–gap plane. The upper plot in Fig. 7 shows a first example, in which the initial conditions were for



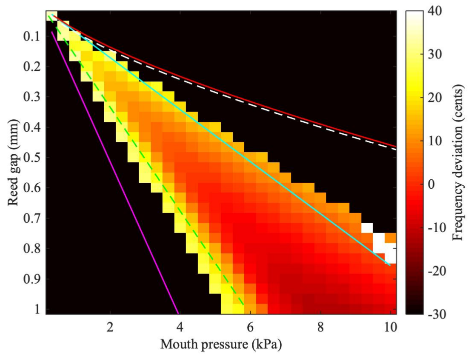


Figure 7. Pressure–gap diagrams for the lowest note of a clarinet, using a "cold start" transient. The upper plot is colour-shaded with the playing frequency normalised by the nominal frequency; the lower plot is shaded to show the frequency deviations in cents from nominal.

a "cold start", with zero pressure inside the tube, and a sudden switching-on of the mouth pressure at time $t = 0$. Black pixels show cases that did not lead to a self-excited note. Coloured pixels show the cases that played a note, coloured to indicate which register was selected.

The colour scale shows the playing frequency (deduced from correlation analysis of the last few period-lengths in each simulation), normalised by the nominal frequency. Nearly all pixels indicate a value close to 1, indicating the lowest register. But a small number of white pixels show cases where the instrument spontaneously jumped to the second register, with frequency approximately a factor of three higher. The lower plot in Fig. 7 shows a different colouring of the same dataset. This time, colours indicate the deviation of playing frequency from the nominal frequency of the note, expressed in cents.

In a striking parallel to the Schelleng and Guettler diagrams, we see a wedge-shaped region within which a note can be produced. If the player wants to blow harder, to make the sound louder, the wedge shape dictates that they must also relax their bite so as to move down a diagonal line in this diagram. In so doing, the lower plot in Fig. 7 shows that the note will tend to play progressively flat, a familiar phenomenon in the clarinet. In this example, the flattening appears to be mainly a consequence of inharmonicity between the tube resonances with this fingering.

Several lines have been included in Fig. 7. These lines represent thresholds of various kinds, from analytic formu-

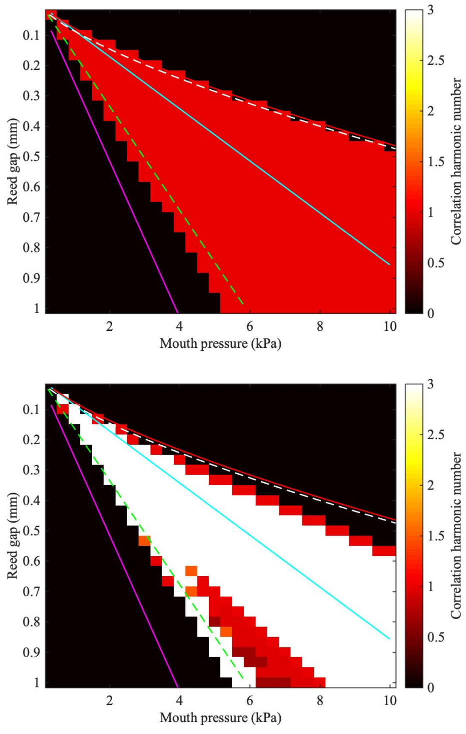


Figure 8. Pressure–gap diagrams for the lowest note of a clarinet as in Fig. 7, but different initial conditions have been used here. In the upper plot, the first tube mode was initialised with a non-zero pressure amplitude. In the lower plot, the second tube mode was initialised with a non-zero pressure amplitude in order to encourage a note in the second register. Both plots are colour-shaded with the playing frequency normalised by the nominal frequency.

lae given by Dalmont et al. [10] on the basis of bifurcation studies. The magenta line relates to the threshold of oscillation: when there is energy dissipation, pixels that spontaneously generate a note from a "cold start" must lie at least a little above this line.

The dashed green line represents the condition Dalmont et al. [10] call the "beating reed threshold". Below this line, the reed remains open throughout the vibration cycle, but above the line the reed is closed for part of the time. The line is based on their analysis, calculated using the clarinet version of Raman's model, but it should give a good guide to the behaviour in a more realistic model.

The solid cyan line represents the condition Dalmont et al. [10] call the "inverse oscillation threshold": it shows when the mouth pressure, for a given gap, is just enough to close the reed completely against the mouthpiece. It is easy to see why this line gives an upper limit in the case of a gentle initial transient, as in Fig. 7: beyond this line, the mouth pressure is high enough that the reed is closed initially, and the gentle transient means that it stays closed for ever after, and no note is produced.

But with a more vigorous transient, the reed has a chance to open after the initial closure. For a pixel lying not too far above the cyan line, it is possible for a note to get going, and then to be sustained. However, for a pixel above the highest (red) line there is simply no possible solution (at least within the simplification of the Raman model, which Dalmont et al. [10] used to derive the condition). This line represents what they call the "extinction threshold". The dashed white line, lying very close to the red line, is what they call the "saturation threshold". This is the condition for the pressure amplitude to be a maximum.

Examples of what might be meant by "more vigorous transients" are shown in Fig. 8. In the upper plot in this figure, each simulation was initialised with a non-zero amplitude of the pressure variation associated with the lowest mode of the tube. This encourages oscillation at a note in the first register, and indeed the plot shows unbroken red pixels now extending all the way up to the extinction threshold.

There is no suggestion that this kind of initial transient is directly relevant to what a player might do. Rather, the intention is to ask a question analogous to the one lying behind the Schelleng diagram. The wedge-shaped region in the Schelleng diagram shows where it should be possible to *maintain* a steady Helmholtz motion, if it can once be initiated. But it says nothing about what kind of transient might be needed for that initiation: for that we need something like the Guettler diagram. So for the clarinet, the red region in the upper plot of Fig. 8 is an attempt to show the maximum region within which it is possible to maintain a first-register note, once it is initiated.

The lower plot in Fig. 8 shows what happens if the *second* tube mode is initialised with non-zero amplitude. This encourages a note in the second register, and indeed many white pixels are now seen. In practice, a similar effect could be produced by starting a note with the register key, then releasing it – often, the second-register note can be sustained by overblowing, without the register key. But the plot shows that some cases chose to play near the first-register frequency, despite the encouragement from the initial condition.

3.3 Results for the soprano saxophone

The clarinet has a tube which is, approximately, cylindrical. But many reed instruments have an approximately conical bore profile, and this has a significant influence on the behaviour. To illustrate, some examples will be shown based on the lowest note of a soprano saxophone. The simulation model is very similar to the one for the clarinet, but this time it is based on a modal fit to the input impedance of a soprano saxophone fingered for its lowest note, taken (by permission) from the website `https://newt.phys.unsw.edu.au/music/saxophone/`.

Figure 9 shows two pressure–gap diagrams, broadly similar to the ones in Figs. 7 and 8. The upper plot shows the results from "cold start" transients. There is a wedge-shaped region quite similar to the one for the clarinet, but this time the note that the model chooses to play from this transient varies within the wedge: first register, second register and third register notes are all in evidence. At least under the assumptions of this simplified model, the saxophone is much more prone than the clarinet to spontaneously overblowing to a higher register.

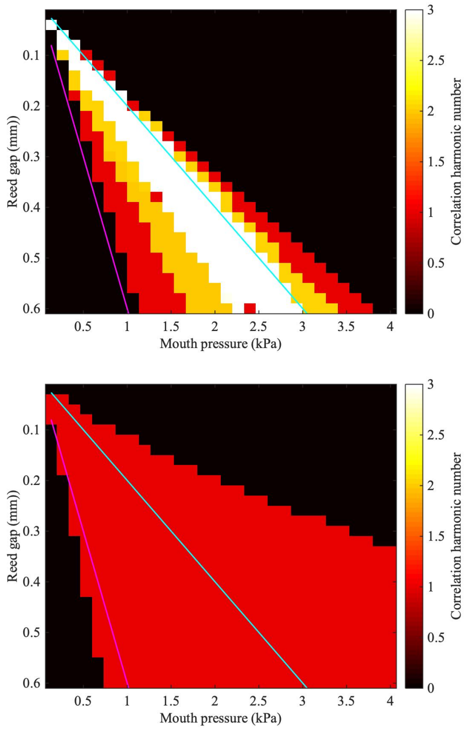


Figure 9. Pressure–gap diagrams for a soprano saxophone fingered for the lowest note. The upper plot uses "cold start" transients, the lower plot is initialised to encourage a note in the first register. Both plots are colour-shaded with the playing frequency normalised by the nominal frequency.

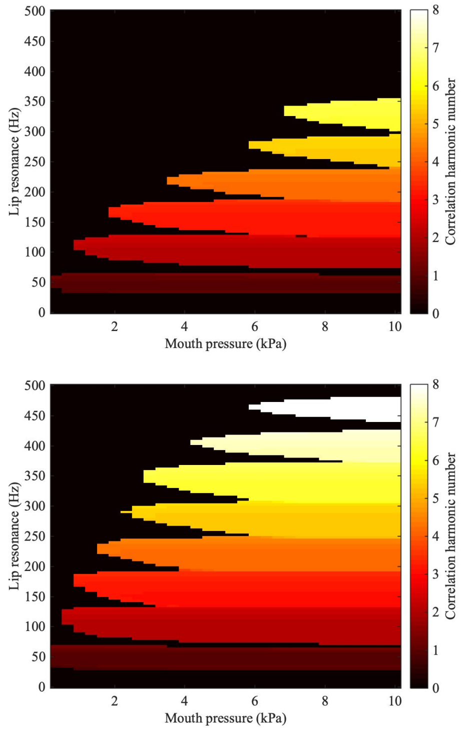


Figure 10. Pressure–gap diagrams for a trombone, using a "cold start" transient. For the upper plot, the Q-factor of the lip resonance is 7; for the lower plot it is 15. Both plots are colour-shaded with the playing frequency normalised by the nominal frequency (defined as the frequency of the second tube mode divided by two).

The lines representing the excitation threshold and the inverse oscillation threshold are included in Fig. 9 because they apply equally well to the saxophone case. But the other lines do not transfer so easily to this case: see section 9.4.8.3 of Chaigne and Kergomard [3] for some relevant discussion.

The lower plot of Fig. 9 shows the effect of "priming" the fundamental mode of the tube, to reveal the maximal area within which a first-register note can be sustained. As with the clarinet, a very large red wedge is revealed. In this case, the bottom edge of the wedge even extends below the excitation threshold. This is evidence of an "inverse bifurcation": there is a range of mouth pressure for which a note can be maintained once started, but cannot be started from low amplitude.

4. BRASS INSTRUMENTS

4.1 The pressure-lip resonance diagram

Brass instruments seem very different from reed instruments, but from the perspective of a simulation model they are surprisingly similar. All that has to be changed in order to convert a reed simulation program into a first attempt at a brass simulation program is a single minus sign. The usual simplified model for the action of a brass-player's lips (see for example Campbell et al. [11]) uses the same equations as the standard model of a reed mouthpiece, except that lips act as an outward-opening reed, not an inward-opening one as on the clarinet or saxophone. Higher blowing pressure tends to open the lips, not to close them. To be sure, the detailed vibration of human lips is certainly more complicated than the vibration of a cane clarinet reed – the single degree of freedom approximation used here is just a first attempt at modelling, in the spirit of Schelleng and Guettler.

The change of sign from inward-opening to outward-opening makes a crucial difference to the behaviour. "Reed" resonance is far more important for lip vibration than it is for a reed instrument. The result of simple linearised stability theory is the prediction that an outward-opening "reed" can only lead to self-oscillation of a tube resonance when the tube resonance frequency lies a little above the reed resonance frequency (see for example section 13.2 of Fletcher and Rossing [12]). That linearised analysis predicts that a brass player must adjust their lip resonance for each note, placing the resonance a little below the desired note.

This immediately suggests a different two-dimensional subspace of control parameters, that might be used to generate a "playability map": the axes will be blowing pressure and lip resonance frequency. Examples will be shown for a trombone model. The measured input impedance of a trombone (provided by Murray Campbell) has been processed by modal analysis in the same way as the earlier reed instrument examples. The lip model uses parameter

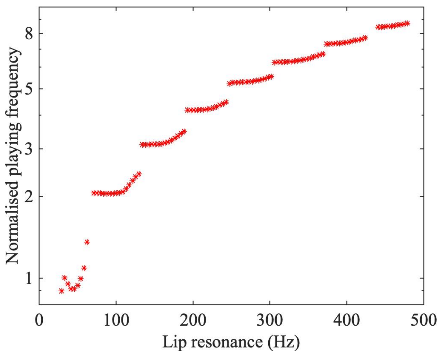


Figure 11. The normalised playing frequency of the trombone model as a function of the lip resonance frequency. Values are taken from the right-most column of the lower plot of Fig. 10. The vertical axis is logarithmic to indicate the relative pitch sensitivity of the different notes.

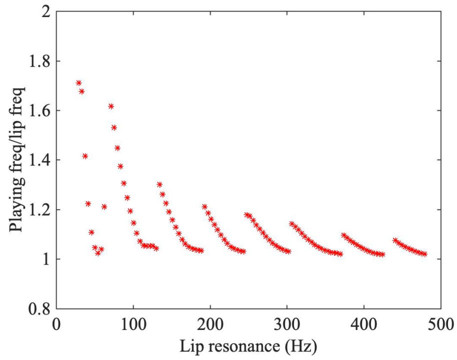


Figure 12. The playing frequency of the trombone model from Fig. 11 divided by the lip resonance frequency.

values taken from Table II of Velut et al. [13].

The upper plot in Fig. 10 shows the result. As the lip resonance frequency is increased, the playing frequency moves up in steps: the first register, then 2nd, 3rd and so on up to the 6th – but no higher with these parameter values. But if the Q-factor of the lip resonance is doubled, the lower plot in Fig. 10 is obtained. The 7th and 8th notes are now seen, and in addition all the threshold pressures for the lower notes have moved to the left. This conclusion is in accordance with the work of Doc et al. [14], who showed that in order to match measured threshold pressures with simulations similar to the ones used here, the Q-factor as well as the frequency of the lip resonance has to rise with the target playing frequency. (However, it must be admitted that the Q-values used here seem implausibly high for a resonance of human lips, suggesting that something important may be missing from the model, and indicating a need for further model development.)

Playing frequencies from the right-most column of the lower plot in Fig. 10 reveal other aspects of how the trombone model behaves. Figure 11 shows this normalised playing frequency plotted against lip resonance frequency, with a logarithmic vertical scale. The plot shows that although each successive regime has a plateau representing an approximate "slotted" pitch, there is considerable variation in pitch. Each note rises significantly in frequency as the lip resonance is increased, before the regime transition to the next note. Especially for the lower notes, these variations in frequency are far too large to be ignored in terms of musical intonation. The player is responsible for the fine tuning of every note, to a far greater extent than in the earlier examples of the clarinet and saxophone.

Figure 12 shows the same data, but this time the playing frequency has been divided by the lip resonance frequency. The plotted symbols all remain above the value 1, confirming the earlier statement that the "reed" resonance always has to be lower than the played frequency.

5. CONCLUSIONS

It has been shown that for a variety of sustained-tone musical instruments, there are interesting two-dimensional subspaces of the player's control variables that lend themselves to graphical representation. The search for such subspaces was motivated by Schelleng's diagram for the bowed string, which has proved valuable over the years by encapsulating an important aspect of the "playability" of a bowed string. Another, quite different, subspace relating to bowed strings gives rise to the Guettler diagram. This captures a different aspect of playability, associated with transients rather than with steady tones.

The idea can be extended to wind instruments. Examples have been presented relating to the clarinet, the soprano saxophone, and the trombone. The resulting diagrams again encapsulate aspects of playability in an immediately graspable form, and perhaps they will prove useful in teaching or research. No doubt other such diagrams remain to be found: the author suggests that the search will prove rewarding.

Another target for future research, in bowed strings and wind instruments, is to use diagrams like these during model development. None of the simulation models used here captures the full, complicated physics of the instruments studied. Further model refinement is called for, and playability maps can give a sensitive way to assess the results. By showing patterns in a subspace of control parameters, these diagrams can give a lot of information about the effects of any particular model enhancement. Some examples of the use of Schelleng's and Guettler's diagrams to assess enhanced bowed-string models can be seen in Galluzzo et al. [7].

Acknowledgments

I thank many friends and colleagues for providing data and helpful discussions on this material: especially Joe Wolfe, Evan Davis, Murray Campbell, Mico Hirschberg, Jean-Pierre Dalmont, Jean Kergomard, Christophe Vergez, Tom Colinot and David Sharp.

6. REFERENCES

[1] C. V. Raman, "On the mechanical theory of vibrations of bowed strings," *Bulletin of the Indian Association for the Cultivation of Science*, vol. 15, pp. 1–158, 1918.

[2] T. Colinot, P. Guillemain, C. Vergez, J.-B. Doc, and P. Sanchez, "Multiple two-step oscillation regimes produced by the alto saxophone," *J. Acoust. Soc. Am.*, vol. 147, pp. 2406–2413, 2020.

[3] A. Chaigne and J. Kergomard, *Acoustics of musical instruments*. Springer-Verlag, 2013.

[4] J. C. Schelleng, "Bowed string and player," *Journal of the Acoustical Society of America*, vol. 53, no. 1, pp. 26–41, 1973. [Online]. Available: ⟨GotoISI⟩://WOS:A1973O714500005

[5] K. Guettler, "On the creation of the Helmholtz motion in bowed strings," *Acta Acustica united with Acustica*, vol. 88, pp. 970–985, 2002.

[6] P. M. Galluzzo and J. Woodhouse, "High-performance bowing machine tests of bowed-string transients," *Acta Acustica united with Acustica*, vol. 100, pp. 139–153, 2014.

[7] P. M. Galluzzo, J. Woodhouse, and H. Mansour, "Assessing friction laws for simulating bowed-string motion," *Acta Acustica united with Acustica*, vol. 103, pp. 1080–1099, 2017.

[8] A. Almeida, D. George, J. Smith, and J. Wolfe, "The clarinet: How blowing pressure, lip force, lip position and reed 'hardness' affect pitch, sound level, and spectrum," *J. Acoust. Soc. Am.*, vol. 134, pp. 2247–2255, 2013.

[9] J.-P. Dalmont and C. Frappé, "Oscillation and extinction thresholds of the clarinet: Comparison of analytical results and experiments," *J. Acoust. Soc. Am.*, vol. 122, pp. 1173–1179, 2007.

[10] J.-P. Dalmont, J. Gilbert, J. Kergomard, and S. Ollivier, "An analytical prediction of the oscillation and extinction thresholds of a clarinet," *J. Acoust. Soc. Am.*, vol. 118, pp. 3294–3305, 2005.

[11] D. M. Campbell, J. Gilbert, and A. Myers, *The science of brass instruments*. Springer-Verlag, 2021.

[12] N. H. Fletcher and T. D. Rossing, *The physics of musical instruments*. Springer-Verlag, 1998.

[13] L. Velut, C. Vergez, J. Gilbert, and M. Djahanbani, "How well can linear stability analysis predict the behaviour of an outward-striking valve brass instrument model?" *Acta Acustica united with Acustica*, vol. 103, pp. 132–148, 2017.

[14] J.-B. Doc, C. Vergez, and J. Hannebicq, "Inverse problem to estimate lips parameters values of outward-striking trumpet model for successive playing registers," *J. Acoust. Soc. Am.*, vol. 153, pp. 168–178, 2023.

A COMPARISON OF FRICTION MODELS FOR BOW-STRING INTERACTION BASED ON EXPERIMENTAL MEASUREMENTS

Ewa Matusiak
Department of Music Acoustics (IWK)
University of Music and Performing Arts Vienna
Anton-von-Webern-Platz 1
1030 Vienna, Austria
matusiak@mdw.ac.at

Vasileios Chatziioannou
Department of Music Acoustics (IWK)
University of Music and Performing Arts Vienna
Anton-von-Webern-Platz 1
1030 Vienna, Austria
chatziioannou@mdw.ac.at

ABSTRACT

Physical models of bowed-string instruments are able to predict many aspects of bow-string interaction. However, reproducing time-domain waveforms of string oscillations is still a difficult problem. Aiming to shed some light on the challenges involved in modelling the frictional interaction between bow and string, this study compares the performance of physical models based on different friction modelling paradigms. To that end, the string is modelled in the absence of any feedback at its end supports. Two types of previously proposed friction models are examined: a static one, where the friction force depends only on the relative velocity between the string and the bow, and a dynamic one, where the relative velocity and the friction force are related through a differential equation. These models are applied assuming single-point bowing, as well as a finite-width bow. Results, based on steady-state measurements carried out on a monochord, show slight differences between the various models. It remains to examine how these differences may amplify in case of full instrument simulations.

1. INTRODUCTION

Bowed-string musical instruments are customarily excited via friction. Several models have been proposed that describe this frictional interaction between the bow and the vibrating string [1–4] that can be used to numerically simulate the oscillations of this coupled system. This study intends to shed some light in the performance of these models, by attempting to numerically resynthesise signals obtained via experimental measurement. The aim is to establish a platform for future work, where additional aspects of the underlying physics of bowed-string instruments will be gradually added to the models and their effect on the string vibrations will be analysed, in comparison to experimental measurements. Such comparisons between measured and simulated time-domain waveforms have been already carried out focusing on different types of instruments, offering additional insight on their function [5, 6].

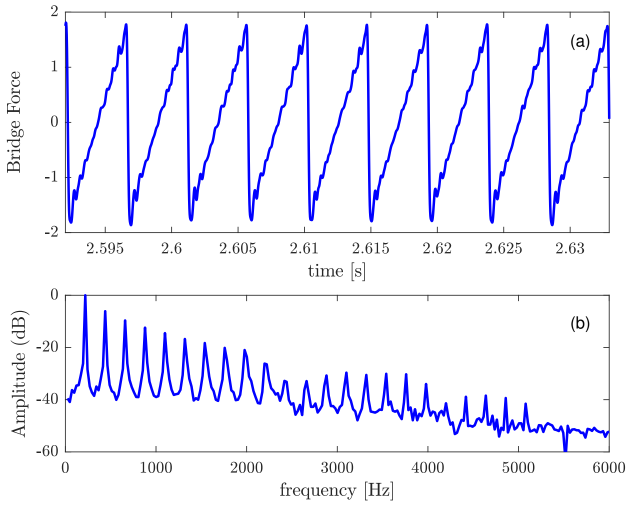


Figure 1. A time snapshot of a measured bridge force signal (a) and its amplitude spectrum (b).

In order to focus only on the bow-string interaction, this study considers the vibration of a single cello string with simply supported boundary conditions. An experimental setup, subject to the same conditions, is used to obtain the target signals for resynthesis (see Fig. 1 for such a signal). This study only investigates the steady-state behaviour of the selected models. The suitability of models based on different paradigms for simulating the frictional interaction between the bow and the string is investigated on the basis of experimental measurements. The insights obtained while trying to resynthesise measured signals for this reduced system should be taken into account, before moving towards the simulation of a full instrument, including transient behaviour.

The friction models considered in this study are described in the next Section, and their numerical discretisation is outlined in Section 3. A comparison between experimentally measured and numerically synthesised signals is presented in Section 4, followed by a discussion of the findings in Section 5.

2. BOW-STRING INTERACTION MODELS

The governing equation for the motion of the string under the bow is the equation describing transverse waves. Using the subscripts t and x to denote differentiation with respect to time and space, respectively, a partial differential equation describing the transverse motion of the bowed string is [7–9],

$$\rho A \partial_t^2 u = T \partial_x^2 u - EI \partial_x^4 u - 2\sigma_0 \rho A \partial_t u + 2\sigma_1 \rho A \partial_t \partial_x^2 u - \mathcal{F}\,, \tag{1}$$

where ρ is the material density, $A = \pi r^2$ the cross-sectional area of the string with radius r, T is the tension of the string, E the Young's modulus, $I = \pi r^4/4$ the area moment of inertia, and σ_0 and σ_1 represent frequency independent and frequency dependent damping. The choice of this relatively simple damping model is motivated by the fact that, at this stage, a small number of fine-tunable parameters is favoured. Future studies may involve more elaborate friction models [10].

Equation (1) is coupled, through the distributed friction force $\mathcal{F}$, to the equation describing the displacement of the bow hair η, namely

$$\mathcal{F} = -K\,\eta - D\,\partial_t \eta\,, \tag{2}$$

where the distributed spring and damping constants are denoted K and D, respectively. In this fashion, the bow is modelled as a simple 'spring-dashpot' [11]. The bow stick is regarded as a rigid frame moving at a given velocity v_B and supporting a ribbon of compliant bow-hair. The relative bow-string velocity is expressed as

$$v = \partial_t u - (v_B - \partial_t \eta)\,, \tag{3}$$

where v_B is the bow velocity. The quantities, u, η, v and $\mathcal{F}$ are functions of time $t > 0$ (in s) and space coordinate $x \in [0, L]$ (in m) for some string of length L. Assuming simply supported ends, the boundary conditions are

$$u(x,t)|_{x=0,L} = 0\,, \qquad \partial_x^2 u|_{x=0,L} = 0 \tag{4}$$

A fourth equation describing the friction force $\mathcal{F}$ is needed to close the system. To this end, paradigms of both static and dynamic nature have been proposed, as outlined in the next section.

2.1 Static Friction Model

The following static friction model assumes that the friction depends only on the relative velocity between the string and the bow, and takes the form

$$\mathcal{F}(v) = f(v) f_N\,, \tag{5}$$

with f_N being the normal force applied by the bow, v being the relative velocity between the bow and the string and $f(v)$ the so-called friction coefficient, a non-linear characteristic which aims to model the stick and slip regimes of friction. The model consists of three components that were developed subsequently and that take care of different aspects of friction. The main idea is that the friction opposes motion, and Coulomb proposed a simple model, velocity independent, saying that the friction force is directly proportional to the normal force, with the friction coefficient being [12]

$$f_C(v) = \mathrm{sign}(v)\mu_C\,, \tag{6}$$

where $\mu_C > 0$ is the so-called dynamic friction coefficient. For lubricated surfaces an additional friction element, viscous friction, was added that is caused by the viscosity of lubricants and is linear in velocity [13],

$$f_V(v) = \mu_V v\,, \tag{7}$$

with $\mu_V > 0$, a viscous friction coefficient. Later, Stribeck observed that during the transition from sticking to sliding the friction drops continuously, and the velocity at which this transition occurs is referred to as Stribeck velocity v_S. Bo and Pavelescu [14] proposed to model the Stribeck effect with a linearised exponential function,

$$f_S(v) = \mathrm{sign}(v)(\mu_S - \mu_C)e^{-(v/v_S)^2}\,, \tag{8}$$

with $\mu_S > \mu_C$ being a static friction coefficient. Finally, the friction coefficient $f(v)$ is a sum of the three components, namely

$$f(v) = f_C(v) + f_S(v) + f_V(v)\,. \tag{9}$$

The above static friction curve has a jump at zero. Therefore, for numerical convenience within a finite-difference time-domain framework, we will use a smooth version of the Coulomb friction curve modelled using the arctan function,

$$\widetilde{f}_C(v) = \mu_C\,\frac{2}{\pi}\mathrm{atan}(v/v_C)\,, \tag{10}$$

where $v_C = 0.02$ and a smooth version of the Stribeck friction curve proposed in [9],

$$\widetilde{f}_S(v) = (\mu_S - \mu_C)\,2\sqrt{a}\,v\,e^{-av^2+0.25}\,, \tag{11}$$

with a a dimensionless parameter. Hence, the overall static friction curve is modelled as

$$f(v) = \widetilde{f}_C(v) + \widetilde{f}_S(v) + \mu_V\,v\,. \tag{12}$$

Fig. 2 shows a smoothed version of the classical friction curve for different values of a.

2.2 Elasto-Plastic Friction Model

As opposed to the static friction model, the elasto-plastic friction model outlined below is a dynamic friction model that relates the relative velocity to the friction force using a differential equation and has been used for simulating the bow-string interaction [15]. The friction force does not depend solely on the relative velocity, but also on an additional variable. In that way, the model exhibits hysteresis loops, a phenomenon observed in measurements using a bowing machine [16]. Moreover, comparing to other friction models proposed in the literature, the elasto-plastic model takes into account presliding behaviour for very small displacements, where the friction increases gradually with the displacement. It also adequately describes friction phenomena at low velocities (i.e. stick-slip motion, presliding behaviour, and, as mentioned above, frictional memory).

The elasto-plastic friction model assumes that the two surfaces, the bow and the string, are irregular at the microscopic level, and their contact is modelled through an

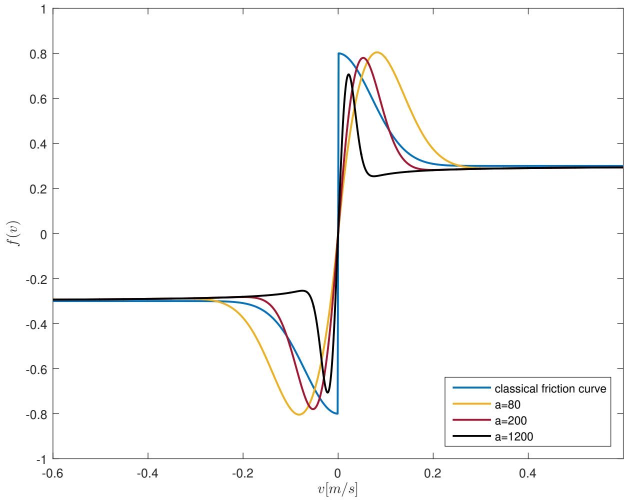


Figure 2. Static friction coefficient curve (blue), and smooth approximations thereof for different values of a. The viscous friction coefficient μ_V is set to zero.

ensemble of elastic bristles, each contributing to the total friction load. The bristles are modelled as damped stiff springs and when the strain exceeds a certain break-away threshold, the bristles break, and the two surfaces begin to slide. Denoting by z the average bristle deflection, the model is given by [3]

$$\mathcal{F}(v, z) = s_0 z + s_1 \dot{z} + s_2 v\,, \tag{13}$$

where s_0 is the bristle stiffness, s_1 is the damping coefficient and s_2 is the viscous friction. Here $\dot{z}$ is the time derivative of z and is related to v through

$$\dot{z} = v\left[1 - \alpha(v, z)\frac{z}{z_{\mathrm{ss}}(v)}\right]\,, \tag{14}$$

where z_{ss} is the steady-state displacement for constant sliding velocities,

$$z_{\mathrm{ss}}(v) = \frac{f_N}{s_0}\Big[f_C(v) + f_S(v)\Big]\,, \tag{15}$$

with f_C and f_S defined in Equation (6) and Equation (8), respectively, and f_N being the normal force applied by the bow. Furthermore, the adhesion map between the bow and the string is defined, for $\mathrm{sign}(v) = \mathrm{sign}(z)$, as

$$\alpha(v, z) = \begin{cases} 0\,, & |z| \le z_{\mathrm{ba}} \\ \alpha_m(v, z)\,, & z_{\mathrm{ba}} < |z| < |z_{\mathrm{ss}}(v)| \\ 1\,, & |z| \ge |z_{\mathrm{ss}}(v)| \end{cases} \tag{16}$$

and $\alpha(v, z) = 0$, for $\mathrm{sign}(v) \neq \mathrm{sign}(z)$. The transition between the elastic and plastic behaviour can be modelled using

$$\alpha_m(v, z) = \frac{1}{2}\Big[1 + \mathrm{sign}(z)\sin\left(\pi\,\theta(v, z)\right)\Big]\,, \tag{17}$$

where

$$\theta(v, z) = \frac{z - \mathrm{sign}(z)\frac{1}{2}(|z_{\mathrm{ss}}(v)| + z_{\mathrm{ba}})}{|z_{\mathrm{ss}}(v)| - z_{\mathrm{ba}}}\,, \tag{18}$$

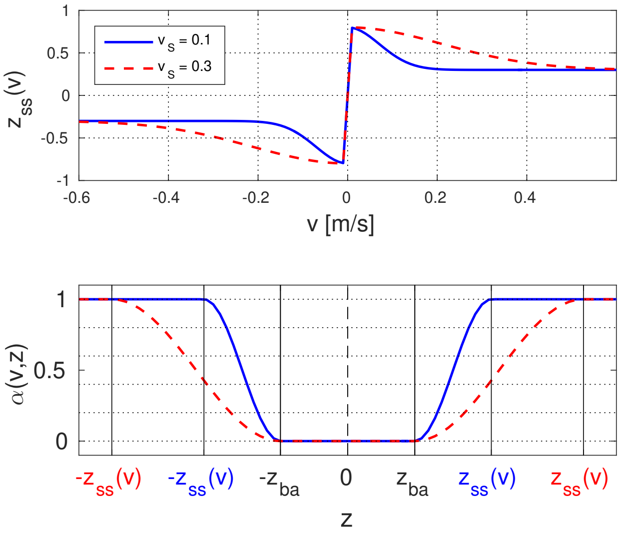


Figure 3. Top: steady-state functions for different Stribeck velocities. Bottom: a curve modelling a bristle displacement z according to the relative velocity of the string with respect to the bow, v, for different steady state functions. When z is between $-z_{\mathrm{ba}}$ and z_{ba} the string sticks to the bow, in the regions $(-z_{\mathrm{ss}}(v), -z_{\mathrm{ba}})$ and $(z_{\mathrm{ba}}, z_{\mathrm{ss}}(v))$ the bristles start to break away (elastic displacement) and when $|z| \ge |z_{\mathrm{ss}}(v)|$ the model enters a plastic phase, and the string slides.

and $z_{\mathrm{ba}} \le \mu_C f_N / s_0$ is the break-away displacement, that is where the bristles start to break. Fig. 3, bottom, depicts the adhesion map for $\mathrm{sign}(v) = \mathrm{sign}(z)$. For small bristle displacements, when $|z| \le z_{\mathrm{ba}}$, $\alpha(v, z) = 0$ and consequently $\dot{z} = v$ (elastic pre-sliding), while for larger displacements, that is when $z_{\mathrm{ba}} < |z| < |z_{\mathrm{ss}}(v)|$, some bristles start to break and a mixed elasto-plastic sliding occurs. Finally, for $|z| \ge |z_{\mathrm{ss}}(v)|$ all bristles break, and a purely plastic regime is achieved, that is the string slips under the bow. In that situation, $\alpha(v, z) = 1$ and the bristle displacement does not change in time, i.e. $\dot{z} = 0$, and consequently the bristles enter a steady state regime, $z = z_{\mathrm{ss}}(v)$.

This elasto-plastic model has been already implemented based on a finite difference method in [17] where it was applied to point-bowing a stiff string. The focus of that article was on real-time implementations and not on reproducing measured signals. A comparison of the elasto-plastic model and a thermal model, introduced in [2], was presented in [15] using a digital waveguide implementation. That analysis concentrated on bringing out differences and similarities between those two dynamic friction models. The numerical stability of a finite-difference implementation of the elasto-plastic model is analysed in [18].

3. DISCRETISATION

A common way to numerically simulate a bowed string is the use of finite-difference time-domain methods. Finite-difference schemes for the string in isolation and the bowed one were described by various authors (see, e.g. [9, 11]). We first discretise the equations of motion, leaving the friction term $\mathcal{F}$ to be discretised at the end. In order to fix the notation, let x_B^L and x_B^R be the positions on the string of the inner and outer bow edges, respectively, with the centre of the bow lying at x_B. Let M be a number of string locations under the bow, denoted by x_m. If $M = 1$ then the model assumes point bowing and $x_M = x_B$.

The model is discretised in time and space with functions u_l^n that are approximations of $u(x,t)$ at positions $(l\Delta_x, n\Delta_t)$. Discretisation in time is performed with $t = n\Delta_t$, where $\Delta_t = 1/f_s$ (in s) with f_s the sampling rate (in Hz) and $n \in \mathbb{N}$, and in space with $x = l\Delta_x$, where the grid spacing Δ_x (in m) has to satisfy the following stability condition [9]

$$\Delta_x \geq \sqrt{\frac{\omega + \sqrt{\omega^2 + 16\kappa^2\Delta_t^2}}{2}}\,, \tag{19}$$

where $\omega = c^2\Delta_t^2 + 4\sigma_1\Delta_t$ with $c = \sqrt{T/\rho A}$ the wave speed and $\kappa = \sqrt{EI/\rho A}$ a stiffness coefficient. The grid points are $l \in \{0, \ldots, N\}$, where $N = \lfloor L/\Delta_x \rfloor$, hence the total number of grid points is $N + 1$.

For a point x_m under the bow, the interpolation operator $I(x_m)$ interpolates the string displacement at position x_m. It is a row vector of size $N+1$ that acts on the column vector $u^n = [u_0^n, \ldots, u_N^n]^T$. On the other hand, a spreading operator $J(x_m)$ is a column vector that distributes the friction force around the bowing point x_m. The two operators are related to each other through

$$J(x_m) = \frac{1}{\Delta_x} I(x_m)^T\,. \tag{20}$$

The approximations of continuous derivatives present in Equation (22) and Equation (24) are the operators

$$\begin{aligned}
\delta_{t+} &= \tfrac{1}{\Delta_t}\left(e_t^+ - 1\right) & \delta_{t-} &= \tfrac{1}{\Delta_t}\left(1 - e_t^-\right)\\
\delta_{t\cdot} &= \tfrac{1}{2\Delta_t}\left(e_t^+ - e_t^-\right) & \delta_{tt} &= \tfrac{2}{\Delta_t}\left(\delta_{t\cdot} - \delta_{t-}\right)\\
\delta_{xx} &= \tfrac{2}{\Delta_x}\left(\delta_{x\cdot} - \delta_{x-}\right) & \delta_{xxxx} &= \delta_{xx}\,\delta_{xx}\,,
\end{aligned} \tag{21}$$

where $e_t^+ u_l^n = u_l^{n+1}$, $e_t^- u_l^n = u_l^{n-1}$, $e_x^+ u_l^n = u_{l+1}^n$, $e_x^+ u_l^n = u_{l-1}^n$ and $1u_l^n = u_l^n$.

To simplify the notation, we first divide Equation (1) by ρA, then discretise it to obtain

$$\begin{aligned}
\delta_{tt}u_l^n &= c^2\delta_{xx}u_l^n - \kappa^2\delta_{xxxx}u_l^n - 2\sigma_0\delta_{t\cdot}u_l^n\\
&\quad + 2\sigma_1\delta_{t-}\delta_{xx}u_l^n - (M\rho A)^{-1}J_B\mathcal{F}^n\,,
\end{aligned} \tag{22}$$

where $J_B = [J(x_1)|\ldots|J(x_M)]$. Assuming simply supported ends, the boundary conditions imply

$$u_0^n = u_N^n = 0\,, \quad u_{-1}^n = -u_1^n\,, \quad u_{N+1}^n = -u_{N-1}^n\,, \tag{23}$$

for all $n \in \mathbb{N}$.

Similarly, we divide Equation (2) by D and discretise it as

$$\delta_{t+}\eta_m^n = -\frac{K}{D}\eta_m^n - \frac{\mathcal{F}^n}{D}\,. \tag{24}$$

Here, for each point x_m under the bow, the bow-hair compliance is computed. Then, assuming that the bow velocity v_B is constant throughout time, the relative velocity at point x_m described in Equation (3) can be discretised as

$$v_m^n = I(x_m)\delta_{t\cdot}u_l^n - (v_B - \delta_{t+}\eta_m^n)\,. \tag{25}$$

We can solve (22) at x_m utilising the expression in (21) for the discrete operator δ_{tt}, as suggested in [9], and using (24), resulting in

$$\frac{I(x_m)J_B\mathcal{F}^n}{M\rho A} + \frac{\mathcal{F}^n}{D} + \left(\frac{2}{\Delta_t} + 2\sigma_0\right)v_m^n + s_m^n = 0\,, \tag{26}$$

where

$$s_m^n = \frac{K}{D}\eta_m^n - I(x_m)\Big(c^2\delta_{xx}u_l^n - \kappa^2\delta_{xxxx}u_l^n \tag{27}$$

$$+ 2\sigma_1\delta_{t-}\delta_{xx}u_l^n + \frac{2}{k}\delta_{t-}u_l^n\Big) + \Big(\frac{2}{\Delta_t} + 2\sigma_0\Big)v_B\,. \tag{28}$$

Depending on the friction model, at each bowing point x_m we need to compute either the relative velocity between the bow and the string v_m^n (static friction model) or the relative velocity v_m^n and the bristle displacement z_m^n (dynamic friction model). For the static friction model, $\mathcal{F}^n$ is discretised as

$$\mathcal{F}^n = f_N f(v_m^n)\,, \tag{29}$$

and we arrive at a system of M equations with M unknowns

$$f_N\left(\sum_{i=1}^{M} H_{m,i}f(v_i^n) + \frac{f(v_m^n)}{D}\right) + \sigma v_m^n + s_m^n = 0 \tag{30}$$

where $H_{m,i} = (M\rho A)^{-1}I(x_m)J(x_i)$ and $\sigma = \frac{2}{\Delta_t} + 2\sigma_0$ that can be solved for v_m^n using the multivariate Newton's method.

When the elasto-plastic friction model is used, (13) is discretised using the finite-difference method following [17] and extended to multiple interaction points under the bow. For each point x_m of the string that is in contact with the bow, we have

$$\mathcal{F}^n = f(v_m^n, z_m^n) = s_0 z_m^n + s_1\dot{z}_m^n + s_2 v_m^n\,, \tag{31}$$

where

$$\dot{z}_m^n = v_m^n\left[1 - \alpha(v_m^n, z_m^n)\frac{z_m^n}{z_{\mathrm{ss}}(v_m^n)}\right]. \tag{32}$$

Substituting $\mathcal{F}^n$ in (26) for the expression in (31), gives us M equations with $2M$ unknowns,

$$\left(\sum_{i=1}^{M} H_{m,i}f(v_i^n, z_i^n) + \frac{f(v_m^n, z_m^n)}{D}\right) + \sigma v_m^n + s_m^n = 0\,. \tag{33}$$

The other M equations are obtained through

$$\dot{z}_m^n - a_m^n = 0\,, \quad \text{with} \quad a_m^{n-1} + a_m^n = 2\delta_{t-}z_m^n\,. \tag{34}$$

4. SIMULATIONS VS MEASURED SIGNALS

A cello A-string was mounted on an optical table, rigidly terminated on both sides. The exact values of the parameters of the string and the bow that we used in the setup are given in Table 1. The values of the spring and damping constants of the bow-hair were taken from the literature [11]. A cello player bowed this monochord at a fixed distance from the 'bridge' (the left end termination of the string). The transversal force exerted by the string on the bridge (henceforth called bridge force) was recorded with a sampling rate of 50 kHz. The recorded signal was filtered with a low-pass Butterworth zero-phase filter of order 6 in order to remove the noise present due to the measurement sensors [19]. Next, a steady-state part of the signal was selected, to which we aimed to fit simulated signals. The simulated bridge force signal is calculated as

$$F^n = \rho A(c^2 \delta_{x\cdot} u_0^n - \kappa^2 \delta_{x+} \delta_{xx} u_0^n)\,, \tag{35}$$

where $\delta_{x\cdot}$ and δ_{x+} are defined analogously to $\delta_{t\cdot}$ and δ_{t+}, with x instead of t in the definitions. The duration of the selected portion was 0.041s, that is 2048 samples. First, we looked at the spectrum of this signal to infer about the approximate bowing position, and concluded that, during the selected time interval, the centre of the bow was positioned at 0.0546m from the bridge, resulting in $\beta = 1/13$. Furthermore, since we are not considering transients in this study, the parameter s_2 in the elasto-plastic model and the parameter μ_V in the static model were ignored. These viscosity-related parameters may significantly affect the length of the transient and need to be considered in future studies.

Throughout the simulations we kept the bow velocity v_B, the normal force f_N and the bowing position β fixed [1], and concentrated on adjusting the parameters related to the friction models in order to fit the bridge force of the simulated signals to the measured one best. In this study, we concentrated in manually fine-tuning the model parameters. This may provide a better intuition regarding how each model parameter affects the simulations. The obtained optimal values of the parameters are listed in Table 2. Note that, since the different friction models also apply an additional damping to the coupled system, the string frequency dependent damping (σ_1) was also adjusted accordingly.

In general, it is possible using any of the tested models to arrive at a decent approximation of the steady-steady part of the measured signals. Comparisons between simulations and measurements are shown in Fig. 4 and Fig. 5. It was observed that the elasto-plastic model adds a significant amount of damping to the system. Therefore, when using this model, the string damping had to be reduced in order to obtain a good match. When the string-bow contact was modelled as a single-point interaction, the amplitude of the 12th harmonic (and in the case of the static model also that of the 11th) is significantly lower compared to the measured signal. This reduction is expected due to the bowing position, with the effect being more prominent in the numerical case for point bowing. In fact, any operation that spreads the bow-string interaction is expected to diminish this phenomenon. This is observed here for the case of a finite-width bow.

[1] Note that, since the cellist was performing on a monochord with no acoustic feedback, the force magnitude might deviate from that on a real instrument.

Parameter	Value	
Length (L)	0.71	[m]
Material density (ρ)	45	[kg/m^3]
Radius (r)	$4 \cdot 10^{-4}$	[m]
Tension (T)	145.13	[N]
Young's modulus (E)	$2.9 \cdot 10^{10}$	[Pa]
Wave speed (c)	308	[m/s]
Fundamental frequency (f_0)	220	[Hz]
Bow-hair width	0.01	[m]
Bow-hair spring const. (K)	$6 \cdot 10^6$	[N/m^2]
Bow-hair damping const. (D)	$3 \cdot 10^3$	[kg/(sm)]

Table 1. Physical parameters of the cello A-string and the bow used in the setup.

Friction		Static		Elasto-Plastic	
Bow		point	FW	point	FW
v_B	[m/s]	0.09	0.09	0.09	0.09
f_N	[N]	0.05	0.05	0.05	0.05
β		0.0769	0.0769	0.0769	0.0769
σ_0	[s^{-1}]	1	1	1	1
σ_1	[m^2/s]	0.005	0.001	0.0005	0.0006
s_0	[N/m]	–	–	$1 \cdot 10^4$	$2 \cdot 10^4$
s_1	[kg/s]	–	–	0.1	0.1
v_S	[m/s]	–	–	0.3	0.3

Table 2. Estimated parameters to approximate the measured bridge force signal for different models (FW stands for finite-width bow).

Due to the intention to keep the model simple in order to avoid fine-tuning too many parameters, torsional waves on the string have not been considered. It can be shown that including them does not change the steady-state bridge waveform considerably (see Fig. 6 (left)). Moreover, in [20] the authors demonstrated that removing torsional waves from the model does not significantly influence the transition to Helmholtz motion. It should be noted, however, that if frequency dependent damping is reduced, then the effect of torsional waves may become noticeable. This is demonstrated in Fig. 6 (right) where σ_1 is reduced by a factor of 100 while no damping is applied to the torsional string motion. Including torsion suppresses the high-frequency oscillations that are observed in the case of a low-damped string.

The Stribeck velocity v_S was taken here to be 0.3, rather than 0.1 as is often used in simulations of bowed-string instruments. Bigger values of the Stribeck velocity change the steady state friction curve z_{ss}, as shown in Fig 3. For higher Stribeck velocities, the steady state friction curve is less steep, meaning that for small variations of relative velocity v, there are small variations in $z_{\text{ss}}(v)$. Hence, the range of z over which combined elastic and plastic presliding occurs does not vary much with changing velocities as opposed to when the Stribeck velocity is low.

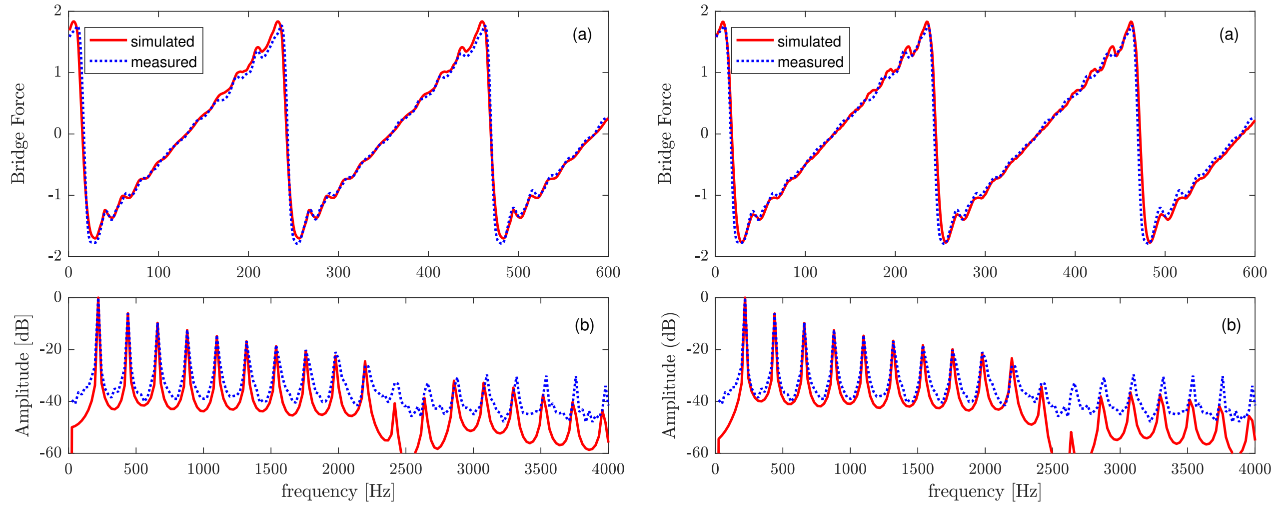


Figure 4. Left: Simulated bridge force for point-bow using the static friction model vs. measurement (a), and their amplitude spectra (b). Right: Simulated bridge force for point-bow using the elasto-plastic model vs. measurement (a), and their amplitude spectra (b).

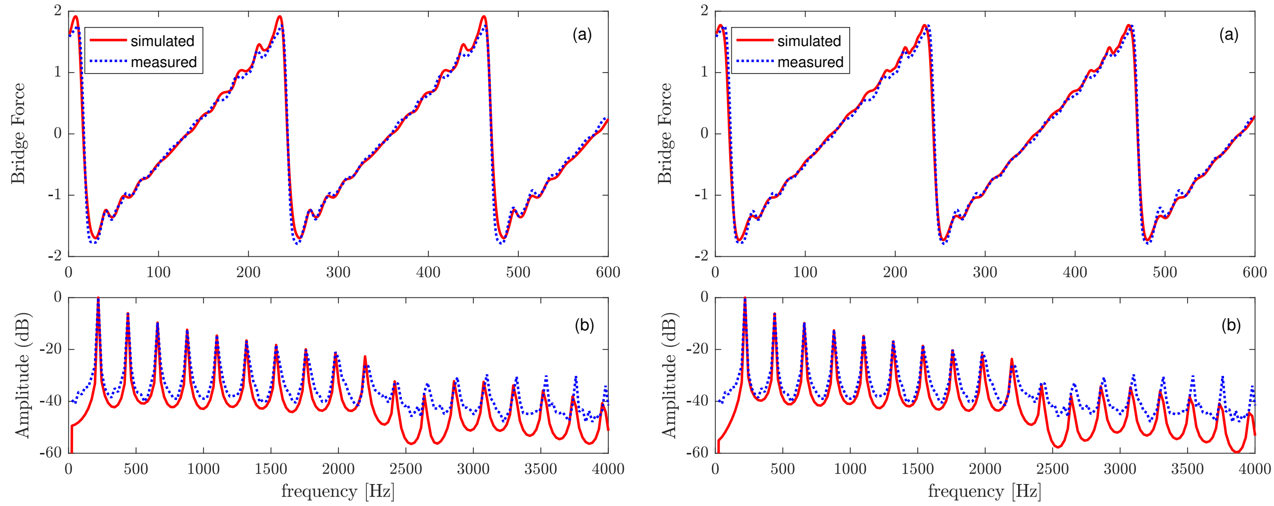


Figure 5. Left: Simulated bridge force for finite-width bow using the static friction model vs. measurement (a), and their amplitude spectra (b). Right: Simulated bridge force for finite-width bow using the elasto-plastic model vs. measurement (a), and their amplitude spectra (b).

5. DISCUSSION

In summary, this study attempts to match an experimentally measured bridge force signal to a simulated one, using previously proposed models. A comparison of different friction modelling paradigms is carried out and some insight is gained on the behaviour of the models, based on manual fine-tuning of the model parameters. Automated parameter estimation, via numerical optimisation, is envisaged. Such a systematic approach should be carried out after a large dataset of experimental signals is collected. This will ideally consist of signals obtained both by human excitation (subject to variations in bow velocity and normal force) as well as by automatic means (e.g. a bowing machine with controlled velocity & force [21]). Furthermore, note transients also need to be considered, in order to obtain a better comparison between the various models and identify whether further model enhancement is required. This study is meant to trigger such future research, while already offering a comparison based on measured steady-state bridge-force signals. After obtaining a good approximation for both transient and steady state-signals for a variety of bowing actions on a monochord, similar methodology could be transferred to a full musical instrument, where the added complexities still evade accurate resynthesis of time-domain waveforms [22].

Acknowledgments

We would like to thank Alexander Mayer and Alessio Lampis for designing the experimental setup and obtaining the measured bridge force signal, as well as Montserrat Pàmies-Vilà and Anna Scheiblauer for several discussions on this subject. This research was funded in whole or in part by the Austrian Science Fund (FWF): P34852-N. For the purpose of Open Access, the authors have applied a CC BY pub-

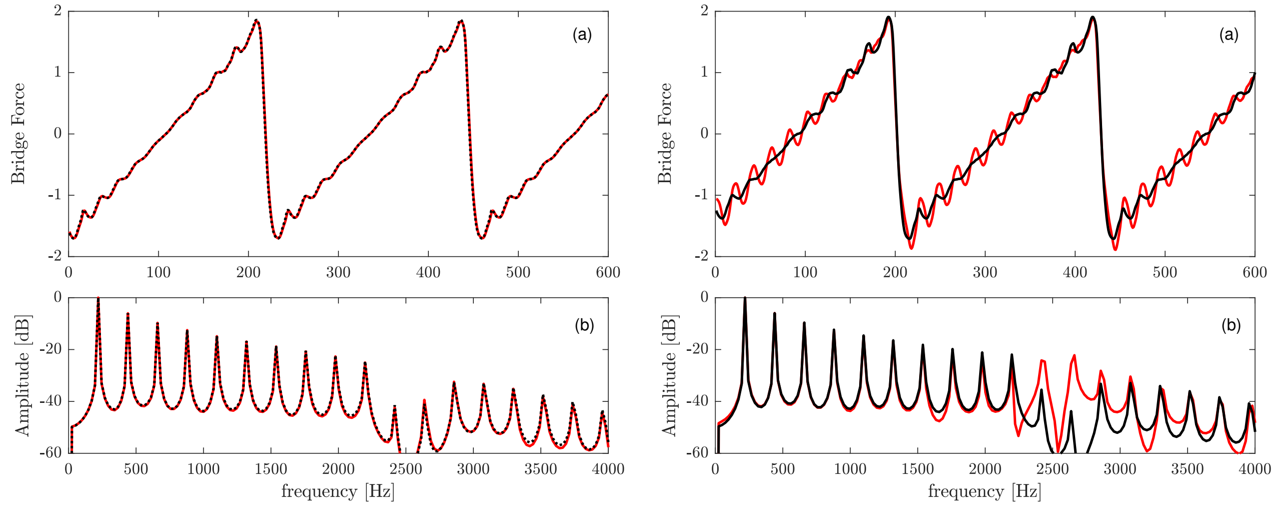


Figure 6. Left: Simulated bridge force for point-bowing using the static friction model with parameters as in Table 2 without torsional waves (red) vs. with torsional waves (black) (a), and their amplitude spectra (b). Right: Simulated bridge force for point-bowing using the static friction model with parameters as in Table 2 with the frequency dependent damping set to $\sigma_1 = 0.00005$ m^2/s. Model without torsional waves (red) vs. a model with torsional waves included (black)(a), and their amplitude spectra (b). In both simulations, the damping factor for torsional waves was set to zero.

lic copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

6. REFERENCES

[1] M. McIntyre and J. Woodhouse, "On the fundamentals of bowed string dynamics," *Acoustica*, vol. 43, no. 2, pp. 93–109, 1979.

[2] J. Smith and J. Woodhouse, "The tribology of rosin," *J. Mech. Phys. Solids*, vol. 48, pp. 1633–1681, 1999.

[3] P. Dupont, V. Hayward, B. Armstrong, and F. Altpeter, "Single state elasto-plastic friction models," *IEEE Trans. Autom. Control*, vol. 47, no. 5, pp. 787–792, 2002.

[4] C. Desvages and S. Bilbao, "Two-polarisation physical model of bowed strings with nonlinear contact and friction forces, and application to gesture-based sound synthesis," *Applied Sciences*, vol. 6, no. 5, pp. 135–167, 2016.

[5] C. Issanchou, J.-L. Le Carrou, C. Touzé, B. Fabre, and O. Doaré, "String/frets contacts in the electric bass sound: Simulations and experiments," *Applied Acoustics*, vol. 129, pp. 217–228, 2018.

[6] V. Chatziioannou, S. Schmutzhard, M. Pàmies-Vilà, and A. Hofmann, "Investigating clarinet articulation using a physical model and an artificial blowing machine," *Acta Acustica united with Acustica*, vol. 105, no. 4, pp. 682–694, 2019.

[7] P. Morse, *Vibration and Sound*. Acoustical Society of America, 1936.

[8] N. Fletcher and T. Rossing, *The Physics of Musical Instruments*. New York: Springer-Verlag, 1991, second Edition: 1998.

[9] S. Bilbao, *Numerical Sound Synthesis*. J. Wiley & Sons, 2009.

[10] C. Desvages, "Physical modelling of the bowed string and applications to sound synthesis," Ph.D. dissertation, The University of Edinburgh, 2018.

[11] R. Pitteroff and J. Woodhouse, "Mechanics of the contact area between a violin bow and a string. Part II: Simulating the bowed string," *Acta Acustica united with Acustica*, vol. 84, pp. 744–757, 1998.

[12] C. Gillmor, *Charles Augustin Coulomb: Physics and Engineering in Eighteen century France*. Princeton, N.J., 1971.

[13] O. Reynolds, "On the theory of lubrication and its application to mr. beauchamp tower's experiments, including an experimental determination of the viscosity of olive oil," *Phil. Trans. Royal Soc.*, vol. 177, pp. 157—-234, 1886.

[14] L. Bo and D. Pavelescu, "The friction–speed relation and its influence on the critical velocity of the stick–slip motion," *Wear*, vol. 82, no. 3, pp. 277–289, 2003.

[15] S. Serafin, F. Avanzini, and D. Rocchesso, "Bowed string simulations using an elasto-plastic friction model," in *Proceedings of the Stockholm Music Acoustics Conference*, 2003.

[16] J. Woodhouse, "Bowed string simulation using a thermal friction model," *Acta Acustica united with Acustica*, vol. 89, pp. 355–368, 2003.

[17] S. Willemsen, S. Bilbao, and S. Serafin, "Real-time implementation of an elasto-plastic friction model applied to stiff strings using finite-difference schemes," in

Proceedings of the International Conference on Digital Audio Effects, 2019, pp. 40–46.

[18] S. Willemsen, "Real-time simulation of musical instruments using finite-difference time-domain methods," Ph.D. dissertation, Aalborg University, 2021.

[19] A. Lampis, A. Mayer, M. Pàmies-Vilà, and V. Chatziioannou, "Examination of the static and dynamic forces at the termination of a bowed string," in *Proc. 184th Meeting of the Acoustical Society of America*, Chicago, 2022.

[20] S. Serafin, J. O. Smith III, and J. Woodhouse, "An investigation of the impact of torsion waves and friction characteristics on the playability of virtual bowed strings," in *Proc. 1999 IEEE Workshop on Appl. of Sig. Proc. to Audio and Acoustics*, New York, 1999.

[21] M. Pàmies-Vilà, A. Scheiblauer, A. Mayer, and V. Chatziioannou, "A framework for the analysis of bowing actions with increased realisticness," in *Proc. 24th International Congress on Acoustics*, Gyeongju, 2022.

[22] P. Galluzzo, J. Woodhouse, and H. Mansour, "Assessing friction laws for simulating bowed-string motion," *Acta Acustica united with Acustica*, vol. 103, no. 6, pp. 1080–1099, 2017.

Sound Analysis of Violin via Neural Network and Numerical Simulation

Masao Yokoyama
Meisei University
masao.yokoyama@meisei-u.ac.jp

Yuki Hori
Bunkyo-Gakki Co.
hori@bunkyo-gakki.com

Shouta Tani
Miyazaki University
hl19051@student.miyazaki-u.ac.jp

Amane Takei
Miyazaki University
takei@cc.miyazaki-u.ac.jp

Genki Yagawa
Toyo University, University of Tokyo
yagawag@gmail.com

ABSTRACT

Two approaches, a neural network and numerical simulation, for analyzing the sound of violins are introduced herein. First, the sounds (open strings and music) of more than 20 violins are recorded. A set of acoustic features, such as the spectrum envelope and mel-frequency cepstrum coefficients, is used to train the neural network and identify the violin timbre. The possibility of quantifying the similarity of the violin timbre and identifying the violin maker is demonstrated. Subsequently, the violins are scanned using a micro-computed tomography scanner to retrieve geometric data, and numerical simulations of the vibration and sound radiation of the violin body are performed. The vibration of the top plate due to the sinusoidal oscillation of the bridge and the sound radiation from the f-hole and C-bouts are calculated using finite-element software COMSOL Multiphysics and ADVENTURE-Sound.

1. INTRODUCTION

Recently, numerical simulations using the finite element method (FEM) have been performed to analyze violin vibrations and sound radiation [1, 2]. To obtain the three-dimensional geometric data of the violin shape, a laser scanner [3] and computed tomography (CT) scanner [4] were used. As shown in these studies, the analysis of violin sounds using computers is flourishing. Numerical simulation is effective for qualitatively and quantitatively modeling vibration and acoustic features. Meanwhile, the timbre of violins is an interesting topic and has been investigated extensively. For example, Setragno [5] and Fritz [6] compared old and modern violins, whereas Corradi et al. [7] and Schleske et al. [8] conducted multidisciplinary analyses of violins.

We conducted a numerical simulation of old violins used by masters, such as the Stradivarius, using the FEM, and modeled the effects of wood properties on the vibration modes of the violin body [9]. Geometric data for the violins were obtained using a micro-CT scanner. Analysis using numerical simulations and CT scanners is a noninvasive and nondestructive method for analyzing historical assets such as Stradivarius violins. Furthermore, we can analyze the sound features of violin using the numerical simulation quantitatively and qualitatively without making lots of violins.

In this study, we recorded the sounds of more than 20 violins, from antiques to brand-new violins. We are currently developing a system that can identify the timbres of violins using neural networks. We attempted to determine whether a computer can identify differences in the timbre of violins as violinists and dealers do. The possibility of identification will contribute to the appraisal and authorization of masters' violins and improve the sound quality of electric instruments and MIDI.

Herein, we introduce our approach for analyzing the sound features of a violin using neural networks and numerical simulations.

2. RECORDING

A list of the violins investigated in this study is shown in Table 1, which includes old Italian violins from the 17th century to the brand-new Japanese commercial violins. We recorded the sound played by a violinist on open strings without any musical expression (E5, A4, D4, G3, longtone of approximately 4 s), the G major scale, and the music piece "Meditation from Thaïs" with musical expression (including vibrato, dynamics, diminuendo, and crescendo).

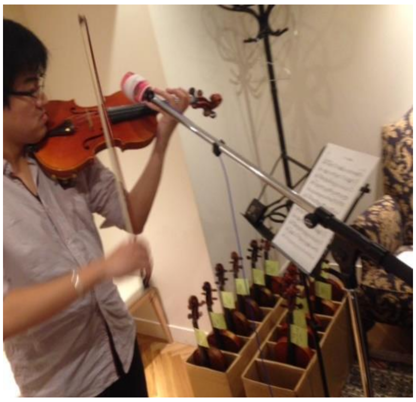


Fig. 1 Recording violin sound using fast Fourier transform analyzer in a rehearsal room.

The piece was played at a slow tempo and had several long notes, and the pitch range was wide, i.e., from A3 (220 Hz) to F#6 (1480 Hz).

A fast Fourier transform analyzer (Oros NV Gate) and a nondirectional condenser microphone (ICP 1/4-inch microphone) were used for recording (Fig. 1). A microphone was placed approximately 10 cm above the violin bridge. Each violin was played twice in a small rehearsal room with reduced echo. The strings of all violins were of the same brand (dominant (A, D, G) and Goldbrokat). Additionally, the same bow and rosin were used.

Violin maker (country)	Year
Catenali (Italy)	ca. 1690
Stradivari (Italy)	1698
Pietro Guarneri (Italy)	ca. 1700
Santo Serafin (Italy)	ca. 1700
Gragnani (Italy)	1760
Balestrieri (Italy)	1780
Pressenda (Italy)	1838
Fabris (Italy)	1870
Scarampella (Italy)	1907
Fagnola A(Italy)	1923
Fagnola B(Italy)	1931
Genovese (Italy)	1927
Michetti (Italy)	1929
Guerra (Italy)	1941
Bisiacchi (Italy)	1953
Garinberti (Italy)	1967
Contemporary violin middle-class (Japan)	2015
Contemporary violin Economic (Japan)	2015
Contemporary violin Stradivari Copy (Japan)	2015
Contemporary violin Guarneri Del Gesu Copy A (Japan)	2015
Contemporary violin Guarneri Del Gesu Copy B (Japan)	2015

Table 1. List of violins used for recording.

3. TIMBRE IDENTIFICATION VIA NEURAL NETWORK

3.1 Identification of timbre based on spectral envelope

The spectral envelope is one of the most important pieces of acoustic information in speech-signal processing. The envelope curve of the power spectrum and local peaks were observed by calculating the power spectrum of the violin sound wave. The peaks emerging in the envelope curve of the spectrum were named the first, second, third, and fourth peaks from lower to higher frequencies (labeled F1–F4, respectively, in this study). Woodhouse [10] mentioned that peaks (humps) over 1 kHz, which are known as transition peaks and bride hill, are similar to vowel formants. For example, the formant frequencies of human

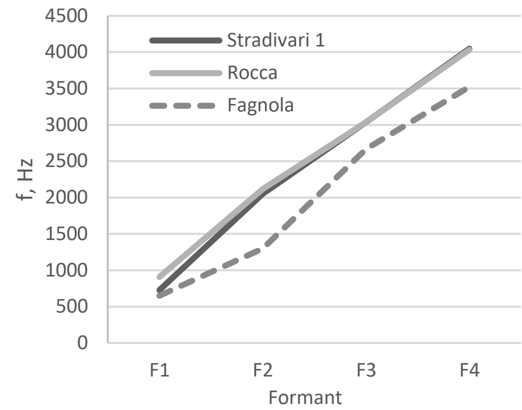


Fig. 2 Distribution patterns of peak frequencies (F1–F4) of power spectrum in open string A[11]. The patterns of the Stradivari and Rocca are similar, whereas that of the Fagnola is different from those of the Stradivari and Rocca.

vowels “a”’ and “e” are different, and we can distinguish an “a” from an “e.”’ Therefore, we applied this property to the identification of violin timbres. We expect that the differences in the timbre can be explained by the patterns of these peak frequencies.

We categorized the peak frequencies (F1–F4) of the recorded sound, as shown in Fig. 2, using the Praat (a sound analysis software) [11]. Some patterns were observed, including linear straight, zigzag, and downward curving patterns. Fig. 2 shows an example of the analysis results, where the patterns of the Stradivari and Rocca violins were almost straight and close to each other, whereas the pattern of the Fagnola violin was zigzagged.

Furthermore, we conducted a perceptual test using a questionnaire to distinguish the timbres of the recorded violin performance [11]. In the test, the participants (10 professional violinists/dealers) listened to three sound waves. After listening to two (A, B) of the three sounds, the participants listened to the third sound (X) and answered whether the timbre of X resembled A or B.

Based on the experimental results, in the case of the sounds of two violins with different distribution patterns, such as straight and zigzag patterns, the participants perceived that the two violins were not similar in terms of timbre. However, when the distribution patterns of the two violins were similar and the difference was insignificant, the participants answered that the two violins were more similar. Thus, the experiment revealed that the shape of the distribution pattern and the difference between distribution patterns can facilitate the identification of violin timbre.

Using the acoustic characteristics of violin sounds, we attempted to determine whether a computer can identify the differences in the timbre of violins. By learning significant amounts of sound data, one may be able to identify certain violins using machine learning techniques.

The acoustic features, i.e., the spectrum envelope and mel-frequency cepstrum coefficients (MFCCs), of the recorded sound data were calculated using the Python software. Neural network training was performed using a pair of acoustic features and violin markers, where the input was the acoustic features and the output was the violin marker. Finally, the trained network was applied to the test data to validate its accuracy and to calculate the similarity for each violin. An identification label was assigned to each of the 21 violins shown in Table 1 and to four strings; thus, 84 labels were assigned to distinguish the timbres.

The neural network was a fully connected four-layer network. The number of inputs was 1024, and in the case where the training data was a spectrum envelope, the number of neurons in the second and third layers was 512. The number of outputs in the fourth layer was the same as that of the labels. The program for calculating the cepstrum as preprocessing and to execute machine learning was written using Python and the Keras library, which is the front end of TensorFlow. The activation function was the ReLU function. The learning rate was 0.1 and the dropout ratio was 0.2. The loss function was a categorical cross-entropy.

3.2 Comparison of accuracies

Fig. 3 shows a comparison of the accuracies of the acoustic features (spectral envelope, mel spectrum, and MFCC). As shown, MFCC is the best option for identifying violin sounds. Additionally, the accuracy was high (approximately 90%) when the neural network was trained and tested using only open-string data. This is because the sound wave of an open string is approximately periodic. However, when a musical piece with musical expression was used as the training data, the accuracy decreased (see middle of the graph in Fig. 3, without string distinction) because the musical piece contained a complex set of parameters, including vibrato and dynamic changes.

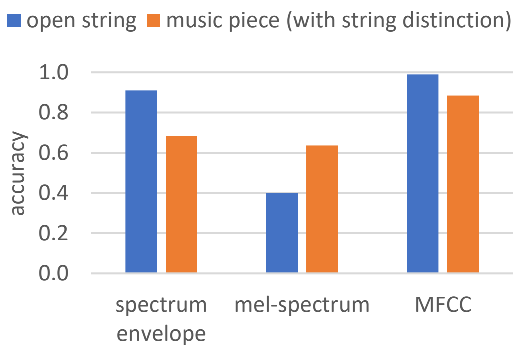


Fig. 3 Accuracy based on nature of sound.

3.3 Similarity in timbre

The probabilities predicted by the proposed neural network are shown in Fig.4. The MFCC data for the four violins selected from Table 1, including Stradivari and Fagnola (shown in legend under the graph) were used to train the neural network. Violins represented on the vertical axis were used to predict the similarity and were assumed to be unknown.

For example, our program predicted that Michetti would be similar to Stradivari by 45%, whereas Gragnani would be similar to P. Guarneri by 83%. However, the timbres of the two band-new violins, the Master-new Del Gesu copies

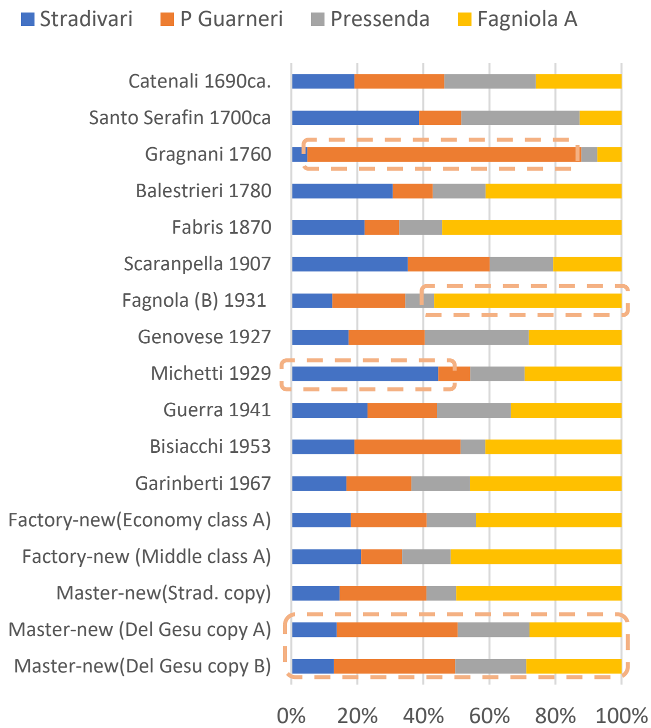


Fig. 4 Percentages of the timbre similarity of test violins against eight famous violins (Dataset: Performance of a music piece, MFCC). Michetti is similar to Stradivari by approximately 45 %. Fagnola B is similar to Fagnola A.

A and B, resembled each other. These two violins are copies of the Guarneri Del Gesu, which was made by the same Japanese master. In addition, Fagnola B was similar to Fagnola A. Interestingly, the sounds of violins from the same maker using the same copy model resembled each other.

Although this result is only an example, we expect that a neural network can indicate the similarity among violins quantitatively and distinguish the timbre of a violin by learning acoustic features, such as the MFCC and spectrum envelope. As an application of this system, when violinists purchase a violin at a shop, the system may provide a useful visualization for quantitatively determining sound features. Violinists who wish to purchase a violin that produces sounds similar to the Stradivari can select one that is the most similar to Stradivari in terms of percentage. In addition, we may be able to develop an appraisal machine using sound to authenticate Stradivari violations in the future.

However, we used only one violin from each violin manufacturer in this experiment. Hence, we cannot definitively conclude that the trained model of the network can express the general characteristics of a violin manufacturer's timbre. More recorded data per violin manufacturer are required for a correct appraisal.

4. SCANNING OF GEOMETRIC DATA USING MICRO-CT SCANNER

The geometries of violins made by Stradivari and Guarneri del Gesu were scanned using a micro-CT scanner, whose precision was 0.1 mm. The upper photograph in Fig. 5 shows a cross-sectional image of the violin body obtained via a micro-CT scan. The lower image in Fig. 5 shows an image of the interior of the violin body (Stradivari, 1719) using scanned geometric data and computer-assisted design (CAD) software.

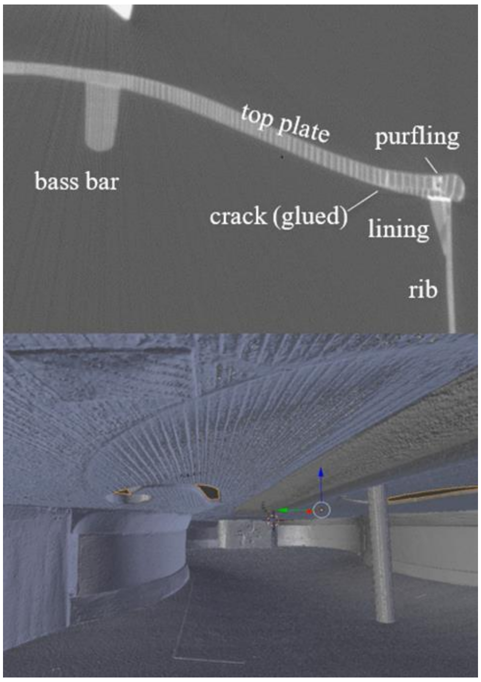


Fig. 5 Visualization of interior of violin via micro-CT scanning (upper side) and CAD software (lower side).

Because the scanned raw geometric data included many fragments and holes, we cleaned the data using the CAD software before performing numerical simulation and visualization. Using this scanner allowed us to observe not only the details inside the violin body, but also the grains of the wood, cracks, and traces of restoration. The geometric data were classified into components such as top and back plates, ribs, sound posts, and bass bars such that different mechanical properties can be specified for each component. The scanned data were saved as STEP files (the standard for exchanging product model data).

5. NUMERICAL SIMULATION OF SOUND RADIATION OF VIOLIN

The import of geometric data, meshing, and FEM calculations were conducted using COMSOL Multiphysics^{TM}[12]. The STEP files created (as mentioned in Section 4) were imported into the COMSOL Multiphysics software as geometric objects (Fig. 6, upper). In addition, a spherical area of air surrounding the violin was constructed (Fig. 6, lower).

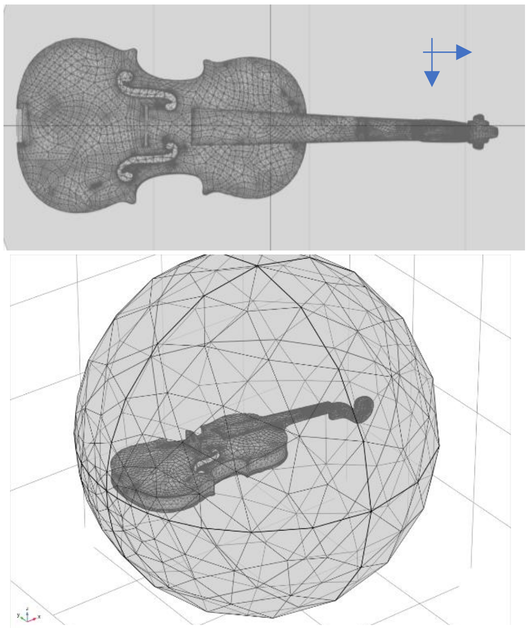


Fig. 6 Mesh of violin and air field generated using auto-mesh function in COMSOL Multiphys-

In COMSOL Multiphysics, the mesh generator discretizes the domains into tetrahedral second-order mesh elements using the free-mesh method. Approximately two million elements were generated, including those for the violin and air. The eigenfrequency, body displacement, and sound pressure were calculated using the FEM via the acoustic–structure interaction module in COMSOL Multiphysics.

The mechanical characteristics of the wood for the violin were set in three orthogonal directions in COMSOL Multiphysics (longitudinal grain direction, radial annual ring direction, and direction tangential to the annual ring). We specified the values for the mechanical properties, such as the Young's modulus, rigidity modulus, and Poisson's ratio, based on measurement values by Green et al. [13]. We set the representative values of density for maple and spruce as 0.63 and 0.36, respectively, and their Young's moduli as 12.6 and 9.9 GPa, respectively.

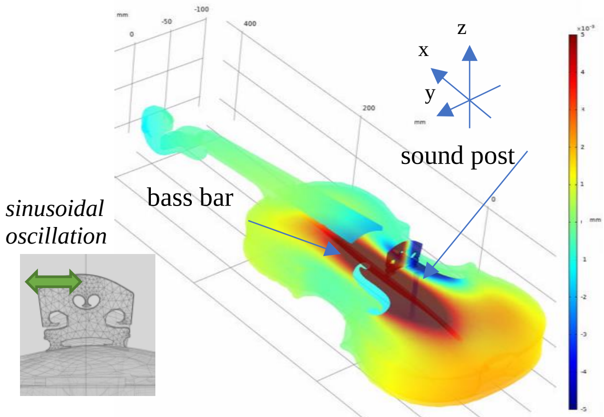


Fig. 7 Displacement of violin by forced vibrations on the bridge, where the G string is placed (196 Hz). Bass-bar side shows significant vibrations. Scroll and fingerboard vibrate as well.

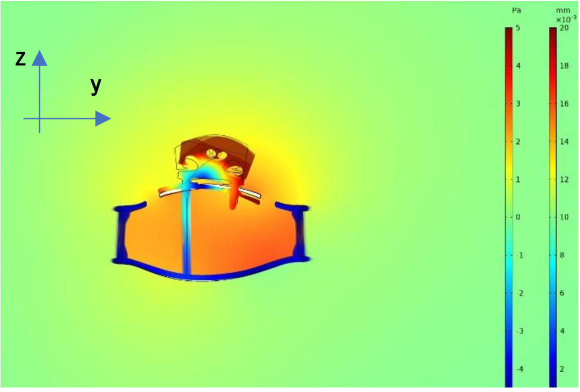


Fig. 8 Visualization of displacement of bridge and sound pressure field in the y–z plane.

5.1 Vibration of violin body

Figure 7 shows the simulation results when a forced sinusoidal oscillation was applied to the bridge of the violin. The color contours depict the displacement of the body in the z-axis direction caused by forced vibrations. The left side of Fig. 7 shows the oscillating position of string G. The frequency of the sinusoidal function at 196 Hz (G3, the fundamental frequency of the G string) was input along the y-axis.

The bridge alternately oscillated from side to side along the y-axis, and alternate vibrations were induced on the top plate by the bridge oscillation. In particular, the magnitude of the displacement on the bass-bar side (left side of the violin) was significant. We speculated that the vibrations of the scroll and fingerboard did not significantly affect the sound volume and timbre; however, these vibrations were not negligible, as shown in the video depicting the detailed simulation results [14].

5.2 Near sound field

The acoustic pressure fields in the y–z plane (including the bridge and sound post) and the x–y plane (30 mm above the arch of the top plate) are shown in Figs. 8 and 9, respectively. Based on the vibration of the top plate caused by the sinusoidal oscillation of the bridge, we simulated the sound radiation in a concentric circle from the f-hole and C-bout. The result obtained was similar to the experimental result obtained by Wang [15], in which sounds at low pitches radiated concentrically from the violin body. In addition, as shown in Fig. 8, the area of the bass bar fluctuated when the sound post was a fulcrum. The sound post functioned as a fulcrum of vibration and a conduit for connecting the vibration between the top plate and backplate.

5.3 Expansion for radiation in a hall using parallel computer

In using the FEM to calculate the sound radiation in a wide area, such as a chamber room or concert hall, billions of meshes are required. Such calculations are difficult to perform on a personal computer; therefore, a large-scale parallel computer is necessitated. We plan to expand our numerical simulation to calculate sound radiation in a concert hall using the parallel computer software, ADVENTURE Sound [16].

The acoustic pressure and coordinate data on the violin surface obtained using COMSOL were transferred to the ADVENTURE Sound software, which is a parallel acoustic analysis code. An iterative domain decomposition method was applied to ADVENTURE Sound. IDDM is the most efficient parallel technique for large-scale analysis, which features several hundred million to several billion degrees of freedom implemented using the hierarchical domain decomposition method [17].

The violin was constrained in a spherical air domain, and the analysis results obtained using ADVENTURE Sound based on a frequency of 196 Hz at the bridge is shown in Fig. 10. The mesh size was 2.76 mm and the number of meshes was approximately 20 million. As a

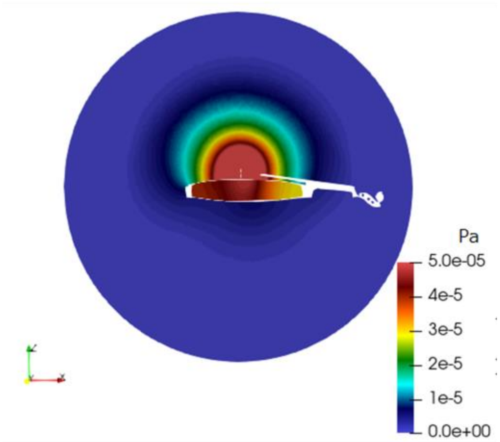
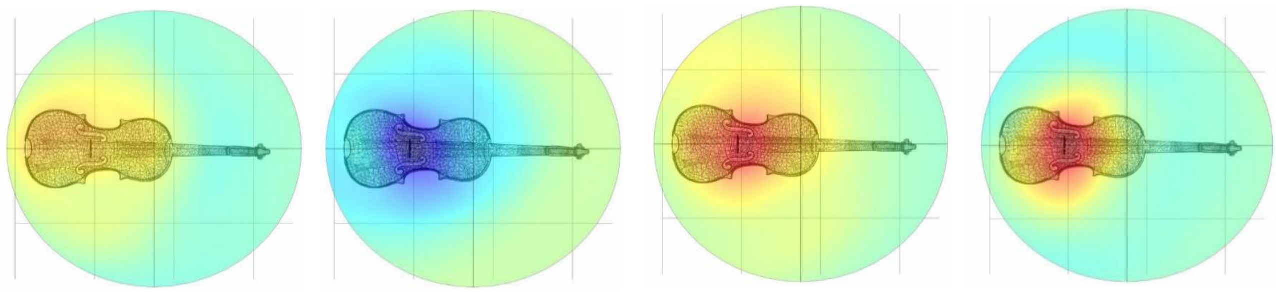


Fig. 10 Visualization of sound radiation around a violin by ADVENTURE Sound.



$t = 0.1T$ ($T = 1/196$ s) $t = 0.15T$ $t = 0.3T$ $t = 0.6T$

Fig. 9 Temporal change in acoustic pressure around violin. T represents a cycle of sinusoidal oscillation on the bridge.

boundary condition, sound absorption was set at the surface of the sphere. A PC cluster (Intel Corei7-9700K, 3.60 GHz, 32 GB RAM/Node, nine nodes) was used for the calculations. The results indicate that the radiation of sounds from the violin surface to the air domain around the violin can be calculated.

Currently, a program for large-scale analysis using acoustic pressure on the surface of a violin body is being implemented. Dynamic analysis of sound radiation from forced vibrations on a violin bridge in a hall will be conducted in the near future.

6. CONCLUSIONS

In this study, experiments using neural networks and numerical simulations were performed to investigate the timbre of violins. We demonstrated the possibility of predicting violin manufacturers based on the acoustic features (spectrum envelope and MFCC) of violins using a deep learning program. To authenticate Stradivari's violin, the sounds of many violins must be recorded; however, an appraisal machine may be realizable using artificial intelligence.

In addition, we scanned old Italian violins using a micro-CT scanner and performed a numerical simulation of sound radiation from the violin body. Experimentally, the sound pressure around an instrument can be analyzed in an anechoic chamber using array microphones. However, this method is expensive, and securing the appropriate facilities for array microphones is difficult. Therefore, a coupled numerical analysis of the vibrations and acoustics can be substituted for the experiments. We are planning to connect these experimental analysis (using recording and neural network) and numerical analysis as future work.

Acknowledgments

This study was supported by JSPS KAKENHI (grant number 21K11969).

7. REFERENCES

[1] C. Gough, "Acoustic characterization of string instruments by internal cavity measurements" in J. Acoustical Society of America, 150(3), 1922-1933, 2021.

[2] J. A. Torres, C.A. Soto, and D. Torres-Torres "Exploring design variations of the Titian Stradivari violin using a finite element model" in J. Acoustical Society of America, 148(3), 1496-1506, 2020.

[3] M. Pezzoli, R.R. De Lucia, F. Antonacci, and A. Sarti, "Predictive simulation of mechanical behavior from 3D laser scans of violin plates" in Proceedings of International Conference on Acoustics 2019, 2019.

[4] G. Bissinger, and D. Oliver, "3-D laser vibrometry on legendary old Italian violins" in Sound and Vibration 41.7, 10-15, 2007.

[5] F. Setragno, M. Zanoni, F. Antonacci and A. Sarti,. "Feature-Based Timbral Characterization of Historical and Modern Violins" in International Symposium on Musical Acoustics, 90-93, 2017.

[6] C. Fritz et al. "Soloist evaluations of six Old Italian and six new violins" in Proceedings of the National Academy of Sciences 111.20, 7224-7229, 2014.

[7] R. Corradi et al. "A multidisciplinary approach to the characteri-zation of bowed string instruments: the musical acoustics lab in Cremona" in Proceedings of the 22nd International Congress on Sound and Vibration(ICSV22), 2015.

[8] M. Schleske, "Empirical tools in contemporary violin making: Part I. Analysis of design, materials, varnish, and normal modes" in Catgut Acoust. Soc. J., 4(5) 50–64, 2002.

[9] M. Yokoyama, "Coupled numerical simulations of the structure and acoustics of a violin body" in J. Acoustical Society of America, 150 (3), 2058–2064, 2019.

[10] J. Woodhouse, "The acoustics of the violin: a review." Reports on Progress in Physics 77.11, 115901, 2014.

[11] M. Yokoyama, "Possibility of distinction of violin timbre by spectral envelope," in Applied Acoustics, 157, 107006, 2020.

[12] Acoustic-Structure Interaction, COMSOL Homepage, https://www.comsol.jp/model/acoustic-structure-interaction-417.

[13] D. W. Green, J. E. Winandy, and E. E. Kretschmann, "Mechanical properties of wood" in Wood handbook: Wood as an engineering material. Madison, WI: USDA Forest Service, Forest Products Laboratory. General technical report FPL; GTR-113: 4.1-4.45, 113, 1999.

[14] https://youtu.be/ m3cgG-TJs9Q

[15] L.M. Wang and C. B. Burroughs, "Acoustic radiation from bowed violins" in Journal of the Acoustical Society of America. 110(1), 543–555, 2001.

[16] ADVENTURE Project home page: https://adventure.sys.t.u-tokyo.ac.jp/ (Access on 15 Nov. 2021) .

[17] A. Kudo and A. Takei, "Performance evaluation of wave-sound analysis code: ADVENTURE Sound" in Proceedings of the 40th JSST Annual Conference: International Conference on Simulation Technology, 1-3, 2021.

Numerical investigation of a plucking position estimation method in the time domain

G O. Paiva
Sorbonne Université, CNRS
Institut Jean Le Rond d'Alembert,
LAM team, Paris, France /
University of São Paulo,
Computer Science Department
São Paulo, Brazil
g.paiva@usp.br

JL Le Carrou
Sorbonne Université, CNRS
Institut Jean Le Rond d'Alembert,
LAM team, Paris, France
jean-loic.le_carrou@sorbonne-universite.fr

ABSTRACT

This paper presents a numerical investigation of a time-domain method for plucking position estimation. The method is based on the time lag assessment between the first and the third pulses passing through at a fixed point close to the bridge, called *measurement point*. This time lag is extracted from the autocorrelation function for one period of the acceleration response signal obtained at the measurement point, which allows the estimation of the plucking position. The influence of the variation of several parameters on the plucking position estimation is studied. By means of a modal synthesis framework a set of pluck simulations is generated using different parameter variations. The proposed method is then applied to estimate the plucking position in each case. In general, the obtained results show good agreement between estimated and actual plucking positions. The plucking position and the natural frequency of the plucked string were found to be the most sensitive parameters. Several limitations of the method concerning the variation of some parameters were also identified. This work focuses on the case of guitar plucks, but it can be promptly extended to other plucked stringed instruments.

1. INTRODUCTION

Despite the numerous studies on the guitar's physics have led to satisfactory sound synthesis results [1–3], a lack of realism is identified in how the guitarists interact with the instrument to produce a sound. Since the musician's gesture in all its complexity is strongly decisive in the sound production of the instrument, a better comprehension of the mechanical parameters that govern the guitar plucking would allow us to control the sound synthesis of guitars in a more realistic way. In this way, every experienced guitarist is able to vary specific gestural parameters related to the plucking action in order to obtain a desirable sound quality. These parameters are the plucking position, the angle with which the string is attacked, the angle with which the string is released, the way of touching the string before the attack (e.g., nail, flesh or plectrum), and the initial displacement of the string, resulting from the plucking force.

Among these parameters, the plucking position plays a major role in the spectral content determination of a guitar tone. For an ideal pluck, string harmonics whose nodes coincide with the plucking position are not excited. This phenomenon leads to the so-called comb-filtering effect on the magnitude spectrum of the string vibration [4]. In practice, this explains why the sound gets brighter as the string is plucked closer to the bridge, since the high-frequency components are enhanced. Moreover, in real playing conditions the plucking mechanism acting on the string, i.e., nail, flesh or plectrum, has a non-zero contact width which acts as a low-pass filter [5].

Several methods have been proposed for estimating the plucking position on a guitar string [6–8]. Traube *et al.* [6] presented a frequency-domain approach in which the error minimisation between the theoretical and measured magnitude spectra of a plucked string allowed the estimation of the plucking position from guitar tones recorded with a microphone. On the other hand, Penttinen *et al.* [8] proposed a time-domain method based on the autocorrelation analysis of guitar tones signals captured by an electromechanical film placed under the saddle of the instrument. The plucking position was calculated from the time delay between the two first pulses arriving at the guitar bridge. This time delay corresponds to the minimum of the autocorrelation function of one period of the signal. Traube *et al.* [7] proposed a hybrid time-frequency domain approach in two stages: (i) similarly to [8], a first approximation for the plucking position was obtained in the time-domain from the autocorrelation function of acoustically recorded guitar tones; (ii) followed by a refinement in the frequency domain using an iterative weighted least-square procedure to minimise the error between the spectral envelope of the measured signal and the theoretical spectral envelope calculated from the first plucking position estimation. Despite allowing reliable results for the plucking position estimation, the above-mentioned methods [6–8] do not use direct measurements of the string vibration.

In this paper, we propose a numerical investigation of a time-domain approach for estimating the plucking position using direct measurements of the string vibration. The studied method is inspired in [8] but uses the string acceleration transients measured on the string (see subsection 2.1) instead of the signals collected by a pickup attached to the bridge. A modal synthesis framework described in Subsection 2.2 is applied to generate a set of simulations of the string pluck responses using different parameter variations and conditions. These results are presented and discussed in Section 3.

2. METHODS

2.1 Plucking position estimation

According to the travelling wave theory, once the string is plucked, two transverse waves travel in opposite directions reflecting back and forth between the string terminations. The interference between the two travelling waves produces a standing wave in the string. The string vibration resulting from the pluck is then attenuated by internal and external damping mechanisms until the string reaches its rest position.

An analysis of the travelling waves propagating through the string just after the pluck provides important information to retrieve the plucking position. For simplicity's sake, the following assumptions are made:

- The two travelling waves in the string are assumed to be ideal impulses.
- The reflection coefficients at the string terminations are equal to -1, which means that no energy losses occur when the travelling waves are reflected at these points.

Fig. 1, which illustrates the temporal evolution of the travelling impulses through a plucked string at four-time instances, t_0, t_1, t_2 and t_3. The string has length L and its ends are fixed at $x = 0$ and $x = L$. x_p and x_m indicate the plucking position and the measurement point locations, respectively. At instant $t = t_0$ the deflection of the string is maximum but it has not yet been released to vibrate freely. At instant $t = t_1$ the impulse that propagates to the right just after the string release reaches the measurement point at $x = x_m$. When $t = t_2$, the same impulse returns to the measurement point after being reflected at the string end at $x = L$. Finally, when $t = t_3$, the impulse that travels initially to the left reaches the measuring point after being reflected in the string end at $x = 0$. It is important to note that the second and third pulses that arrive at the measurement point at $t = t_2$ and $t = t_3$, respectively, will have opposite phases to the first pulse that arrives at $t = t_1$.

Hence, from Fig 1, the following kinematic expressions can be written:

$$t_1 - t_0 = \frac{x_m - x_p}{c} \tag{1}$$

and

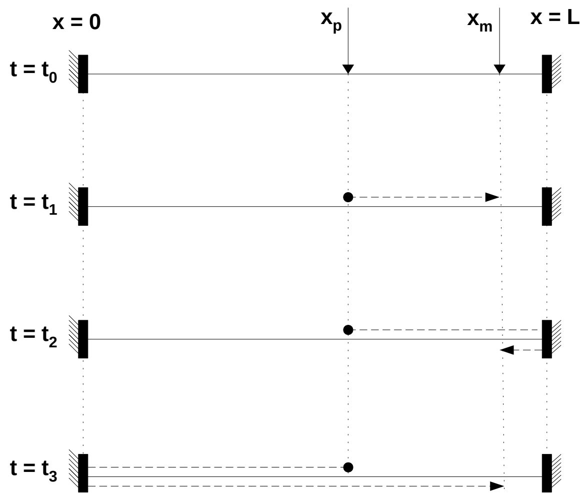


Figure 1. Representation scheme describing the travelling waves in a plucked string at the time instances t_0, t_1, t_2 and t_3. x_p and x_m indicate the plucking position and the measurement point locations, respectively.

$$t_3 - t_0 = \frac{x_m + x_p}{c}, \tag{2}$$

where $c = \lambda f_0$ is the velocity of the transverse wave. f_0 and λ are the string's fundamental frequency and its respective wavelength. By subtracting Equation 2 from Equation 1 and considering $\lambda = 2L$ one can obtain

$$x_p = \Delta\tau L f_0, \tag{3}$$

where $\Delta\tau = t_3 - t_1$ is the time lag between the instants at which the first and the third pulses arrive at the measurement point x_m. Thus, the plucking point can be estimated by determining $\Delta\tau$. In terms of digital signal processing, this time difference is expressed by $\Delta T = \Delta\tau f_s$, where f_s is the signal sampling frequency, which yields

$$x_p = \frac{\Delta T}{f_s} L f_0. \tag{4}$$

In order to determine the time lag ΔT, the autocorrelation function (ACF) [9] of the first period of the acceleration signal collected at the measurement point is then computed. Two negative peaks are expected in the ACF since the second and third pulses are anti-symmetric with respect to the first. Thus, the time lag at the second negative peak of the computed ACF corresponds to ΔT.

Fig. 2a and Fig. 2 show two typical examples of signals obtained by modal synthesis (see Section 2.2) when the string is plucked at $x_p = 0.05$ m and $x_p = 0.60$ m, respectively. In those figures, the time instances t_1, t_2 and t_3 are identified in both waveforms and the phase opposition between the first and the third pulses can be verified. In both cases, the acceleration response is obtained at $x_m = 0.01$ m and the values for the other parameters are $L = 0.65$ m, $f_0 = 196$ Hz, $f_s = 44.1$ kHz. Fig. 2b and Fig. 2d display the corresponding ACFs for the signals shown in Fig. 2a and Fig. 2c, respectively. For both cases, the second negative peaks are observed and indicated by red stars.

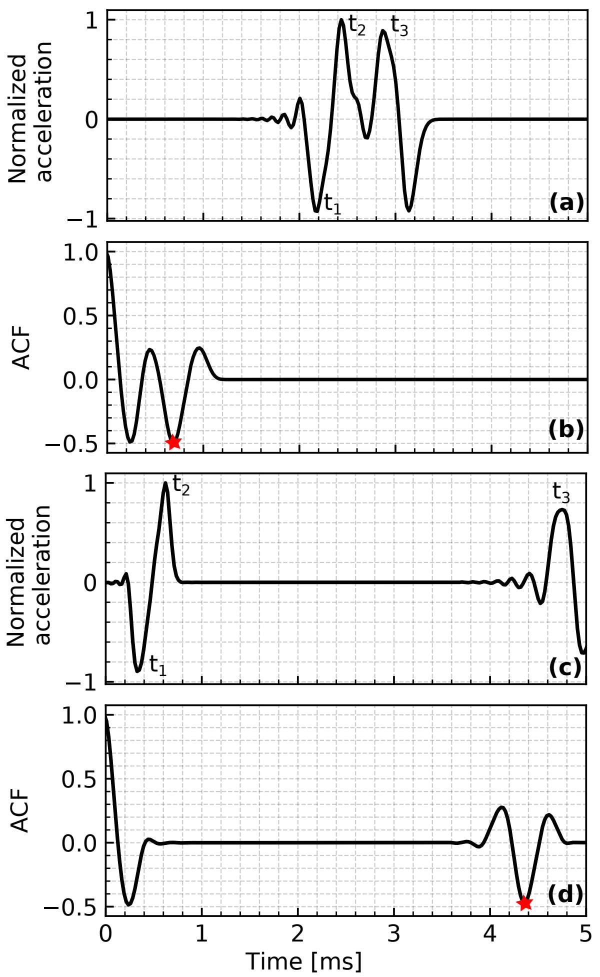


Figure 2. (a) First period of a synthesised acceleration signal when the string is plucked at $x_p = 0.05$ m (b) and its respective autocorrelation function. (c) First period of a synthesised acceleration signal when the string is plucked at $x_p = 0.60$ m (d) and its respective autocorrelation function. In both cases, the acceleration responses were measured at $x_m = 0.01$ m. The red stars indicate the second local minima of the ACFs.

In a first stage, the second negative peak value is obtained by determining the second local minimum value of the ACF. In practice, an empirically determined threshold is initially applied to filter small peaks and eventual noise. The first and the second negative peaks are then identified by comparison of neighbouring values. In a second stage, the accuracy of the identification is improved by fitting a second-order polynomial to the initially estimated value of the second minimum and its two neighbouring left and right values. This refinement procedure allows for fractional values of ΔT. In order to analyse the robustness of the plucking position estimation method, noise is added to the synthesized signals (see Subsection 2.2). In these cases, a Savitzky–Golay filter [10] is applied to the noisy signal and to the computed ACF for the purpose of smoothing the data.

2.2 Numerical string signal

A modal framework is used to synthesise multiple string plucks, which allows obtaining the acceleration responses at the measurement point x_m. The plucking position estimation method is tested in these signals. The adopted synthesis model is fully described in [11]. Basically, this approach uses the string and the instrument body modal parameters to synthesise the string transient response resulting from a pluck. Mode shapes, natural frequencies and damping factors of the uncoupled strings are obtained from analytical expressions while the modal basis of the instrument body is extracted from an experimental modal analysis [11]. All the string parameters necessary to obtain the synthesised results presented in this work were extracted from [2], while all the body parameters were extracted from the experimental modal analysis reported in [11]. It is assumed that the string and instrument body motion have a single polarisation, parallel to the plucking direction. The string and instrument body are coupled by imposing displacement continuity at the coupling point located at the saddle. A temporal finite difference scheme is used to compute, at each instant, the string pluck response in terms of its acceleration at the measurement point x_m. This synthesis method also allows to simulate the response of an isolated string, without coupling to the body. In these cases, the mobility of the instrument body was set to zero.

The plucking force is modelled similarly to [1]: a simplified "stick-slip" mechanism simulating the string/finger interaction during the pluck is considered. The string is pulled for 15 ms and released during 0.4 ms. This force is assumed to be exerted at a single plucking point x_p, except in the cases where the contact width is considered. The measurement point is fixed at $x_m = 0.01$m for all the simulations presented in this work.

When designing an experiment, some parameters can be sensitive so their variation may affect the reliability of the plucking point estimation. Therefore, the variation of two sets of parameters will be tested: the *measurement parameters* set, which is defined by the experimental designer; and the *instrument parameters* set, which is intrinsic to the musical instrument or to the musician. These parameters are varied as follows:

- Measurement parameters:
  - *Measurement noise*: signals without measurement noise are initially considered. A signal-to-noise ratio (SNR) of 10dB is then added to the data;
  - *Measurement sampling frequency*: the signal sampling frequency is varied from 18 kHz to 380 kHz.
- Instrument parameters:
  - *String's stiffness*: synthesised signals with and without considering the string stiffness are obtained. The plucking point estimation method is tested in both signals;

- *Number of string modes*: the number of string modes included in the signal synthesis is varied from 25 to 145;
- *Coupling to the instrument body*: a set of simulations considering a single string coupled to a Brazilian guitar (also called *Viola caipira*) body is obtained. The experimental modal parameters of the Brazilian guitar body used in these simulations were extracted from [11];
- *String's fundamental frequency*: different classical guitar strings were also considered in the simulations. The strings and respective fundamental frequencies are as follows: E (82.41 Hz), A (110 Hz), D (146.83 Hz), G (196 Hz), B (246.94 Hz) and e (329.63 Hz);
- *String excitation*: the contact width between the excitation mechanism and the string is varied from 0.1 to 2.5 cm.

The influence of these parameters will be studied according to the estimation error indicator, which is defined as follows:

$$Error = x_p - x_0, \tag{5}$$

where x_p is the plucking position value found by the algorithm and x_0 is the actual plucking position value used in the simulation.

3. RESULTS AND DISCUSSIONS

Multiple simulations varying the studied parameters are performed in order to obtain synthesised signals at the measurement point. For each synthesised signal, the algorithm detailed in the Subsection 2.1 is applied and the estimated result is compared to the actual plucking position used for the simulation. The error criterion can thus be calculated from Eq. 5.

A summary of the results is shown in Fig. 3. Fig. 3a shows the estimation errors when the sampling frequency is varied. Fig. 3b shows the estimation errors when the number of string modes is varied with and without considering the string stiffness. For both Fig. 3a and Fig. 3b, the plucking position is fixed at 0.10 m. Fig. 3c shows the estimation errors obtained when the coupling (or its absence) is considered with different plucking positions. Fig. 3d shows the errors for a variable x_p obtained by adding SNR and changing the string fundamental frequencies. It turns out that for the width of the excitation, the error in the plucking position estimation is found to be equal to 0.2 mm for values ranging from 0.1 cm to 2.5 cm for a perfectly flexible string, for a sampling frequency of 44.1 kHz and for a plucking position at 0.10 m. Similarly, the plucking position estimation seems independent of the coupling with the soundboard (see Fig. 3c). From Fig. 3b, one can note that the variation of the number of string modes has no influence on the plucking position estimation. From the same figure, it is observed that the inclusion of the string's stiffness increases the estimation error: for a flexible string a systematic error of -0.40 mm is obtained, while for a stiff string the error jumps to -1.70 mm.

The other parameters have a more direct impact on the plucking position estimation. In particular, the sampling frequency must be greater than 200 KHz for the error on the estimate to be less than 0.2 mm. However, the lowest sampling frequency tested, 17.5 kHz, leads to an estimation error of approximately 0.25 mm, which shows that the method does not require a high sampling frequency. Noise, on the other hand, introduces a random error to the result as shown in Fig. 3d. According to Fig. 3c and Fig 3d, the main source of error in plucking point estimation is its location itself relative to the measurement point x_m. Indeed, it can be observed that even the string tension changes, the error remains more significant for positions closer to the measurement point. Thus, the plucking position x_p seems to be the most sensitive parameter. This limitation is clearly due to the temporal nature of the method. Despite this, the method is found to be, in most cases, particularly relevant to estimate the plucking point with an error of less than 5 mm.

Finally, some questions remain regarding the applicability of the presented method in an experimental context. In practice, measurements of vibration on a guitar string may be difficult to obtain. To deal with possible difficulties, the authors intend to use measurement techniques based on opto-switch sensors [12] and high-speed camera setup. These methods have been applied in previous works for studying harpsichord [13] and harp [14, 15] plucks, and seem promising if adapted for the guitar case.

4. CONCLUSIONS

A numerical investigation of a plucking position estimation method has been presented. This method works in the time domain and is based on the time lag assessment between the first and the third pulses passing through at a fixed point close to the bridge. A modal synthesis framework was used to generate string plucks with different parameter variations. The proposed method for estimating the plucking position was then applied to these generated signals, which allowed us to investigate the influence of the parameter variation.

The contact width associated with the excitation mechanism and the coupling to the instrument body seems to have no influence on the plucking position estimation. Choosing a sufficiently high value for the sampling frequency is an important factor in obtaining estimates with relatively low errors. The following parameters have a significant impact on the plucking position estimation: the added noise (SNR), the string fundamental frequency, and the actual plucking position x_p, which seems to be the most influential.

For experimental validation purposes, the proposed method will be applied to signals measured in different string musical instruments, in different musical contexts. The comparison of the presented method with other previously proposed methods is also planned for future work.

Acknowledgments

The authors are thankful to the São Paulo Research Foundation (FAPESP), grant no 2022/07224-0, and to French National Center for Scientific Research (CNRS) for funding.

5. REFERENCES

[1] G. Derveaux, A. Chaigne, P. Joly, and E. Bécache, "Numerical simulation of a guitar," *Computers & structures*, vol. 83, no. 2-3, pp. 107–126, 2005.

[2] J. Woodhouse, "Plucked guitar transients: Comparison of measurements and synthesis," *Acta Acustica united with Acustica*, vol. 90, no. 5, pp. 945–965, 2004.

[3] S. Bilbao and A. Torin, "Numerical modeling and sound synthesis for articulated string/fretboard interactions," *Journal of the Audio Engineering Society*, vol. 63, no. 5, pp. 336–347, 2015.

[4] N. H. Fletcher and T. D. Rossing, *The physics of musical instruments*. Springer Science & Business Media, 2012.

[5] A. Chaigne and J. Kergomard, *Acoustics of musical instruments*. Springer, 2016.

[6] C. Traube and J. O. Smith, "Estimating the plucking point on a guitar string," in *Proc. Conf. on Digital Audio Effects (DAFX'00), Verona, Italy*, 2000, pp. 153–158.

[7] C. Traube and P. Depalle, "Extraction of the excitation point location on a string using weighted least-square estimation of a comb filter delay," in *Proceedings of the Conference on Digital Audio Effects (DAFx)*. Citeseer, 2003, pp. 188–191.

[8] H. Penttinen and V. Välimäki, "A time-domain approach to estimating the plucking point of guitar tones obtained with an under-saddle pickup," *Applied Acoustics*, vol. 65, no. 12, pp. 1207–1220, 2004.

[9] L. Rabiner, "On the use of autocorrelation analysis for pitch detection," *IEEE transactions on acoustics, speech, and signal processing*, vol. 25, no. 1, pp. 24–33, 1977.

[10] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures." *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[11] G. O. Paiva, F. Ablitzer, F. Gautier, and J. M. C. dos Santos, "Collisions in double string plucked instruments: Physical modelling and sound synthesis of the viola caipira," *Journal of Sound and Vibration*, vol. 443, pp. 178–197, 2019.

[12] J.-L. Le Carrou, D. Chadefaux, L. Seydoux, and B. Fabre, "A low-cost high-precision measurement method of string motion," *Journal of Sound and Vibration*, vol. 333, no. 17, pp. 3881–3888, 2014.

[13] A. Paté, J.-L. Le Carrou, A. Givois, and A. Roy, "Influence of plectrum shape and jack velocity on the sound of the harpsichord: An experimental study," *The Journal of the Acoustical Society of America*, vol. 141, no. 3, pp. 1523–1534, 2017.

[14] M. Pàmies-Vilà, "Caractérisation du pincement de la harpe de concert en contexte musical," Master's thesis, Pierre and Marie Curie University, 2015.

[15] D. Chadefaux, J.-L. Le Carrou, B. Fabre, and L. Daudet, "Experimentally based description of harp plucking," *The Journal of the Acoustical Society of America*, vol. 131, no. 1, pp. 844–855, 2012.

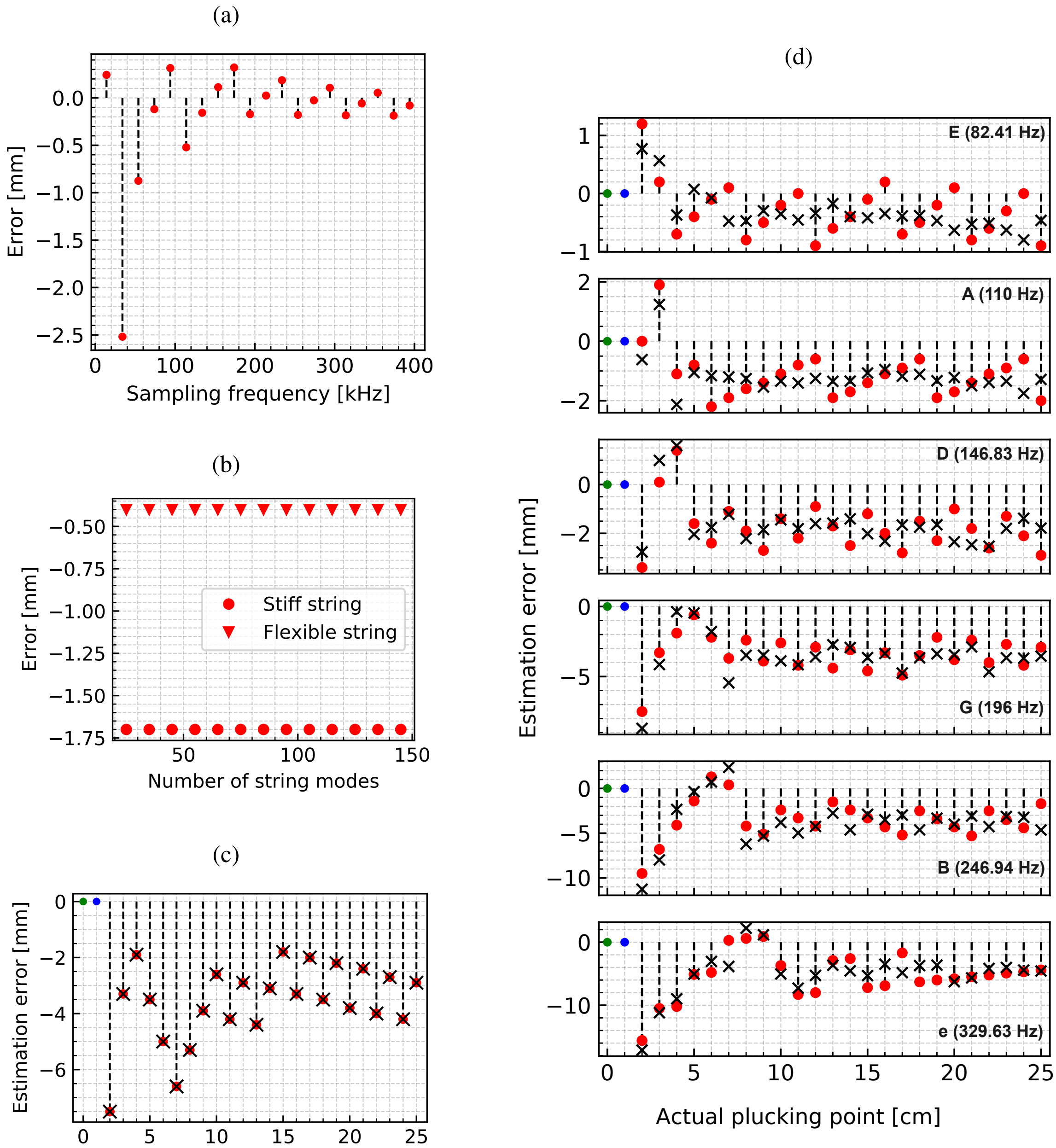


Figure 3. Error of the plucking position estimation obtained by the algorithm as a function of (a) the sampling frequency, (b) number of string modes and string stiffness, (c) the string-soundboard coupling (red circles and black crosses are the estimated plucking position without and with coupling), (d) fundamental frequency and SNR (red circles and black crosses are the estimated plucking positions with and without noise, respectively). For (a) and (b), an isolated string, without coupling, is simulated. For (a), (b) and (c) a G-string (f_0 = 196 Hz) is synthesised. For (c) and (d), the green circle is the string-soundboard coupling point; the blue circle is the measurement point. For (d), each subplot shows the result for each guitar's string.

Effect of Soundpost Position Adjustments on the Perceptual and Vibrational Qualities of the Violin

Ailin Zhang
Shenzhen University
az304@szu.edu.cn

Lei Fu
Luoyang Institute of Science and Technology
bjfulei.good@163.com

ABSTRACT

The soundpost is commonly used by bowed string instrument makers as an important link between the plates. Research into the function of the soundpost is of necessity and significance. The goal of this paper is to gather the variations in both vibration modes and human perception in response to changes in the soundpost position. An adjustable soundpost and a performance level violin were used in the experiment. The soundpost was shifted towards or away from the centerline (left or right), closer or further from the bridge (upward or downward) by 5 mm separately during the experiment. Physical measurements of vibration modes of the violin with the soundpost in different positions were made using an input hammer and a laser vibrometer. Eight experienced violinists and violin or bow makers were invited to play and evaluate the violin. They were asked to compare the violin with soundpost in each of the adjusted positions and with soundpost in the original position and rate them on continuous scales according to seven criteria. Perceptual results showed no significant differences except that the violin became significantly softer or more balanced with soundpost position moved upwards or right respectively compared to the soundpost in the original position.

1. INTRODUCTION

The soundpost, which plays a significant role in the sound making process of bowed-string instruments, is normally a small wooden cylinder. Not only does it support the top plate of the instrument, but also influences the vibration modes of plates. The tone of the instrument might change dramatically due to the alternation of the soundpost. By moving the position of a soundpost, the bowed-string instrument can make a louder, clearer, darker, or brighter sound. Thus Musicians usually have their soundpost adjusted by luthiers before important concerts, especially if the instrument has a muffled, unclear or fuzzy sound.

The traditional soundpost of the violin family is normally made by wood. The length of a soundpost in a stringed instrument like a violin or cello is determined by the arching of plates, the size of the instrument, and the preferences of the player. In a violin, the soundpost with the diameter about 6 mm is normally located about 2.5 mm below the treble foot of the bridge, between the back and front plates of the instrument. The soundpost should be perfectly perpendicular to the contact planes of top and back plates. Note that there is no glue or adhesive to be applied: the soundpost is only held in position by compression. Thus the exact length and the position of the violin soundpost is adjusted to optimize the tone of the instrument.

To set up a soundpost in a violin, a special setter is used to insert and put the soundpost into the ideal position through the f-holes without opening the instrument. The soundpost should be moved by gently pushing it to different locations, which is best done by a skilled luthier. Then the violin maker re-tightening the strings to secure the soundpost in place. This process might take a few hours and have a significant impact on the timber of the violin.

Therefore the traditional wooden soundpost on a violin has following shortcomings: firstly, only a few professionals such as violin makers or string instrument repair person can do the adjustment; secondly, materials for making the soundpost vary from spruce to American pine, which might affect the sound radiation; thirdly, the grains of the spruce in the plates are easily damaged since the contacting areas between the soundpost and plates bear a considerable amount of force under full string pressure.

The occurrence of composite materials might bring the adjustment of a violin soundpost the new method. New designed soundpost made by composite metals can be moved easier and used to explore the influence from the position of a soundpost on the vibration of violin plates. This paper examines the influence of a new-designed height adjustable soundpost on the vibration properties and perceptual qualities of the violin and correlate the acoustical behaviour and the perceptions from players. Following the literature review presented in Section 2, the experimental procedures and results are then provided in Section 3 and 4. Discussions and conclusions are given in the final section.

2. LITERATURE REVIEW

In 1970's, Jansson [1,2] began to study the influence from the soundpost on the timber and vibration modes of a violin by hologram interferometry. Using some artificial and immovable wooden soundposts, he concluded that the timber of a violin might be more sharper while moving the position of the soundpost upward, more softer while moving the soundpost downward, more darker while moving the soundpost left, and more lighter while moving the soundpost right. In 1996, Saldner, Molin, and Jansson [3] found that a nonsymmetric mode with the soundpost of a violin is a combination of one symmetric mode and one asymmetric mode without the soundpost.

During the 1990's, Bissinger [4] analyzed the accelerance spectra of a violin with and without the soundpost. The correlations between 12 important vibration modes ranging from 200 to 1000 Hz with and without the soundpost were calculated.

In 2010's, Nadarajah and Woodhouse [5] built a model of the violin body and soundpost and provide an insight into the effects of both tightness and fitness of a soundpost on its vibrational behaviour. Gough [6] calculated the influence on the vibrational modes of a centrally placed soundpost. He claimed that the force from the string to the instrument body can only excite symmetric modes of the plates if the soundpost is set in a central place.

In 2021, Fu, Fritz and Scavone [7] explored how the height of a soundpost affects the perceptual qualities of the violin. Their perception experiments showed that there is no consistent view of the relative optimal soundpost height for the players or makers.

3. EXPERIMENTS

3.1 Acoustical Measurements

In the acoustic measurements, the vibrational modes of top and back plates of a violin were tested with a traditional wooden soundpost and a new-designed soundpost made by composite metals fitting in different positions.

3.1.1 Set Up

The tested violin is made by a Chinese violin maker who has medalled at several international violin making competitions. The strings of the violin were tuned optimally and damped during the measurements. Other set-up details were adjusted properly prior to the experiment.

The original soundpost of the tested object is a small cylindrical piece of wood made by spruce with weight of 0.54 g. The original position of the wooden soundpost is located about 2.5 mm below the treble foot of the bridge as shown in Figure 1. The distance between the f-hole and the treble foot is about 16.5 mm.

The new-designed soundpost made by composite metals with weight of 2.75 g as shown in Figure 2. It is composed of a cylinder shell, two bearing joints, a special core, and a Pythagoras wheel. Comparing with a wooden soundpost, this new adjustable one can change its height much more easier by rotating its Pythagoras wheel. Therefore the alternation of the soundpost position can be adjusted easily and the installation time can be shortened.

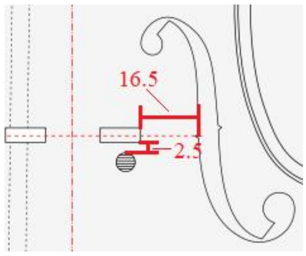


Figure 1. The position of the traditional wooden soundpost on the tested violin.

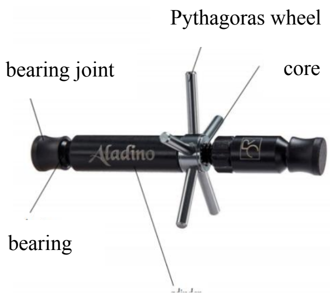


Figure 2. A new-designed adjustable soundpost.

The acoustic measurements on the tested violin supported by soft foams were carried out with fitting the wooden soundpost and adjustable soundpost to different points as shown in Figure 3. An impact hammer (PCB 086C80) provided the input force to the G-string corner of the bridge and a laser vibrometer (Polytec PSV-500) was used to measure the velocity responses of the top and back plates. The laser signals were collected in the direction perpendicular to the surface of the plates at 113 fairly spaced points on the top plate and 88 points on the back one. For each point, the measurements were carried for three times.

Figure 3. Measurements using the hammer and the laser vibrometer on the tested violin.

3.1.2 Measurements

During the experiments, the violin was investigated with two soundposts fitting in following different conditions:

Condition	Soundpost	Position
1	Wooden soundpost	Original position
2	Adjustable soundpost	Original position
3	Adjustable soundpost	5 mm left
4	Adjustable soundpost	5 mm right
5	Adjustable soundpost	5 mm upwards
6	Adjustable soundpost	5 mm downwards
7	No soundpost	NA

Table 1. The tested violin with different soundposts placing in seven different conditions.

The wooden soundpost was only fitted in its original position since its length is incapable of alternation. For the purpose of contrast analysis, the adjustable soundpost was first putted into the same original position. Then the adjustable soundpost was shifted to four different positions: 5 mm left (toward the centerline from its original position), 5 mm right, 5 mm upwards (closer to the bridge), and 5 mm downwards. Usually, luthiers adjusted the soundpost position in the range of 2-3 mm, larger distance beyond this range might damage the violin as the arc of the plates varies. By making use of a height-adjustable soundpost as shown in Figure 2 in this experiment, 5 mm soundpost position adjustment is possible. 5 mm is big enough so that the possible variations of the violin can be attributed to the influence of the soundpost position adjustment other than minor uncontrollable factors. The soundpost adjustment during this experiment was made by a luthier, who could adjust the soundpost position with an accuracy of about ±1 mm. And the luthier varied the height of the soundpost to make the soundpost tension all about the same for any of the experiment condition. The violin was also measured without any soundpost. Note that the pitch of the violin without soundpost was tuned a little bit lower than that of normal case.

3.2 Perception Experiments

The aim of the perceptual experiments is to explore how changes in soundpost position affect the perceptual qualities of the violin. The same violin and adjustable soundpost were employed as in the vibration measurements.

3.2.1 General Design

A luthier and two experimenters were present to adjust the soundpost position and facilitate the experiment. Figure 5 shows the luthier in the experiment.

The experiment included four phases, and there was one pair of playing trials in each phase: one of the playing trials always referred to the violin with soundpost in the original position; the other one corresponded to one of the four adjusted positions. During each phase, each subject was provided with a sheet of rating paper. During each trial, each subject played the violin and wrote his/her general feelings about the violin or variations comparing with the previous trial of the same phase, and rated the violin on continuous scales according to seven criteria. Each subject rated the violin on the same rating scales for the two playing trials during each phase, thus the rating distances between the two trials implied the perceptual differences of the violin with the two different soundpost positions.

Figure 5. Violin maker in the perception experiments.

Figure 6. Player in the perception experiments.

3.2.2 Venues and Controls

The perceptual experiment took place in two cities for the convenience of participants. Five subjects participated in the experiment in Shanghai, and three in Shenzhen. The location in Shanghai is a normal hotel room. The venue in Shenzhen is a music rehearsal room. The rooms both had relatively dry acoustics which can avoid coloring the direct sound from the instrument too much by room reflections.

Subjects used a bow that was provided by the experimenters while assessing the violin. Figure 6 shows a subject evaluating the violin.

3.2.3 Participants

Eight subjects participated in this experiment, among them there were 4 violinists, 3 violin makers and 1 bow maker. They were all native Chinese speakers. Violinist participants had at least 12 years of playing experience (average years of playing=18), two of them had bachelor degrees in music performance, and the other two were currently undergraduate student in music performance. The three violin makers had minimum 5 years of experience in violin making (average years of making=13). The bow maker had been making string instrument bow for 17 years.

3.2.4 Detailed Procedure

Subjects were scheduled to come together at the same time for the experiment in any of the two cities. The experiment lasted around 2 hours.

The experiment consisted of 4 phases as shown in Table 2, and each phase included one pair of playing trials.

Phase	First trial	Second trial
1	Original position	5 mm right
2	5 mm upwards	Original position
3	5 mm left	Original position
4	Original position	5 mm downwards

Table 2: Four phases in the perception experiments.

The soundpost position presentation order was arranged for both the convenience of the luthier and avoiding keeping the original position in the first (or second) trial for all phases.

Before the experiment, subjects were asked to complete a questionnaire and signed the consent form. All forms used in the perceptual experiment were printed in Chinese. The luthier stayed at a room beside the experiment room to adjust the soundpost position. Two experimenters were present in the experiment room for instructing the subjects and pass the violin between subjects and the luthier. It took about 5-10 minutes for the luthier to adjust the soundpost position as designated. Thus there were 5-10 minutes break between trials and phases.

All subjects sat at the experiment room through the whole experiment. During the first trial of each phase, each subject played the violin for several minutes successively. Once finish one's playing, the subject was asked to write down his/her impression about the violin on the provided sheet of paper, and rated the violin on continuous scales according to seven criteria: softness, balance, richness, brightness, responsiveness, resonance and preference. The criteria were carefully chosen from previous publications [2,8] as well as a pilot study conducted before the experiment. The definitions of some of the criteria were explained to subjects [8] so that they can interpret them in consistent ways. Subjects were asked not to communicate their answers during the experiment. After all subjects accomplish the first trial of the first phase, one of the experimenters took the violin to the luthier for soundpost position adjustment. Subjects were told that the luthier may made some adjustments of the soundpost or not. After the adjustment, each subject started the second playing trial of the phase. Subject was required to write down his/her impression or variation about the violin comparing with the first trial. Then subject rated the violin on the same rating scales as the previous trial. In this way, the rating distance on each scale indicated the subject's perceptual difference between the two trials. The violin was brought to the luthier for soundpost adjustment again after the first phase. The procedure was repeated for the other three phases.

4. RESULTS

This section explores the influences from different soundpost positions on the vibration of the violin and preference of the players.

4.1 Results of Acoustical Measurements

As to the limitations of space, this paper only studies one important vibration mode of the top and back plate of the tested violin---Helmholtz resonance mode of the top and back plate of the tested violin.

Helmholtz resonance mode, which is also called as "A0 mode" or "air mode", just as its name implies, is a vibration mode involving air flow in and out of the f-holes. We will explore the air modes of the violin with two soundposts fitting in seven different conditions as mentioned in Table 1.

4.1.1 Condition 1

When the traditional wooden soundpost is placed in its original position, Helmholtz resonance mode of the tested violin appears obviously with the plates taking in and expelling air as shown in Figure 7 and 8. The vibration of the top and back plates show that how does the air volume inside the body of the instrument change at the natural frequency of air mode. The top plate of the tested violin is efficiently driven by the bridge and the soundpost around 262-271 Hz. The two sides of the island area in the top plate are moving out of phase. The vibration in the side opposite to the soundpost is larger than the soundpost side.

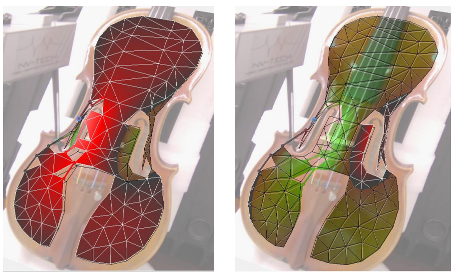


Figure 7. Helmholtz resonance mode of the top plate of the tested violin with a wooden soundpost around 262-271 Hz.

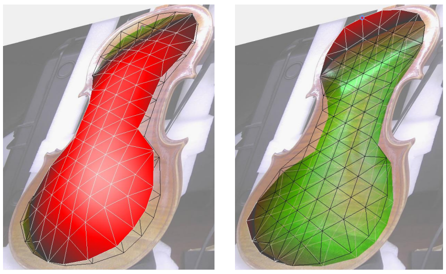


Figure 8. Helmholtz resonance mode of the back plate of the tested violin with a wooden soundpost around 262-271 Hz.

4.1.2 Condition 2

As shown in Figure 9 and 10, after settling the adjustable soundpost at the same position in the tested violin instead of a wooden one, the air mode around 256-271 Hz is very similar to that with a traditional soundpost. But its dB level is a little bit lower than the previous one.

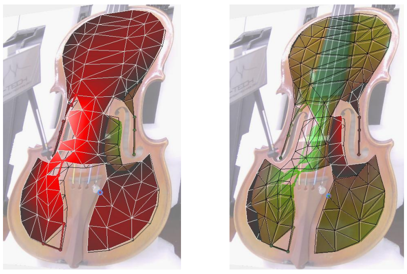


Figure 9. Helmholtz resonance mode of the top plate of the tested violin with an adjustable soundpost around 256-271 Hz.

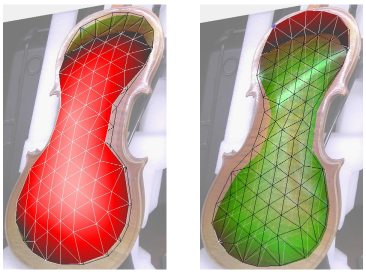


Figure 10. Helmholtz resonance mode of the back plate of the tested violin with an adjustable soundpost around 256-271 Hz.

4.1.3 Condition 3 and 4

While the adjustable soundpost is shifted 5 mm left (toward the centerline form its original position) or 5 mm right, the vibration distributions around A0 mode look alike: the plates are twisting rather than breathing as shown in Figure 11, 12, 13 and 14. Note that under these two conditions the vibrations around A0 resonance might not be a pure mode. The two sides of the island area in the top plates are still moving out of phase. But the amplitudes of vibration in the top and back plates are both smaller. Note that some vibrating areas on the back plates are suppressed.

The mode frequencies under these two conditions are remarkably higher than previous results with the adjustable soundpost in the original position. The dB levels are almost the same.

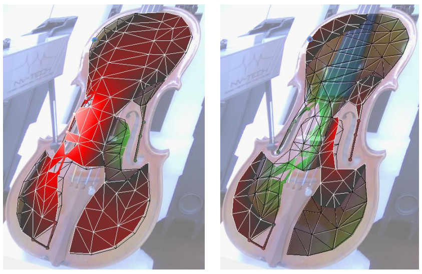


Figure 11. Helmholtz resonance mode of the top plate of the tested violin with an adjustable soundpost shifting 5mm left around 268-275 Hz.

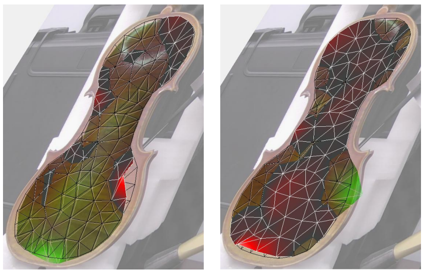


Figure 12. Helmholtz resonance mode of the top plate of the tested violin with an adjustable soundpost shifting 5mm left around 268-275 Hz.

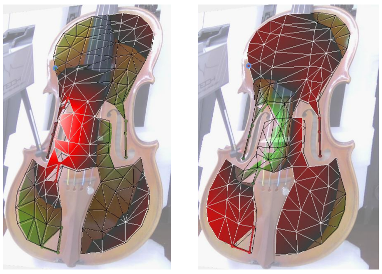


Figure 13. Helmholtz resonance mode of the top plate of the tested violin with an adjustable soundpost shifting 5mm right around 265-281 Hz.

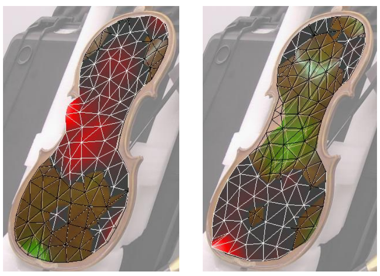


Figure 14. Helmholtz resonance mode of the top plate of the tested violin with an adjustable soundpost shifting 5mm right around 265-281 Hz.

4.1.4 Condition 5 and 6

With the adjustable soundpost moved 5 mm upwards, the air mode around 253-256 Hz is similar to that with an adjustable soundpost in the original position. The two sides of the island area in the top plate are moving out of phase as shown in Figure 15. However, the back plate seems like a longitudinal dipole as shown in Figure 16.

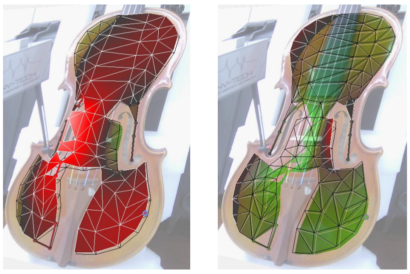


Figure 15. Helmholtz resonance mode of the top plate of the tested violin with an adjustable soundpost shifting 5mm upward around 253-256 Hz.

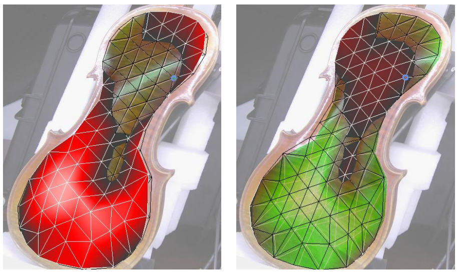


Figure 16. Helmholtz resonance mode of the top plate of the tested violin with an adjustable soundpost shifting 5mm upward around 253-256 Hz.

With the adjustable soundpost moved 5 mm downwards, the plates are breathing air around 259-262 Hz as shown in Figure 17 and 18. Its air mode shows high resemblance with that in condition 2.

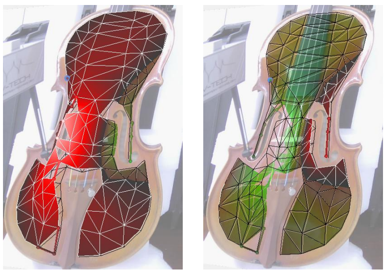


Figure 17. Helmholtz resonance mode of the top plate of the tested violin with an adjustable soundpost shifting 5mm downward around 259-262 Hz.

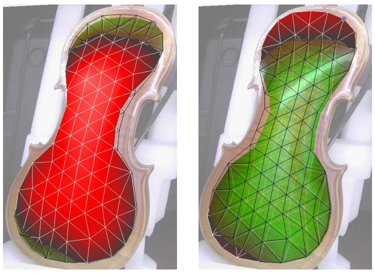


Figure 18. Helmholtz resonance mode of the top plate of the tested violin with an adjustable soundpost shifting 5mm downward around 259-262 Hz.

4.1.5 Condition 7

After removing the soundpost, all the resonance frequencies are decreased. The vibration modes seem more symmetrically distributed without any soundpost, as shown in Figure 19. The vibration in the side opposite to the soundpost is almost the same with that in the soundpost side although the two sides of the island area in the top plate are still moving out of phase. That could be because the soundpost gives an additional fixed point between plates and suppressed the vibration near it.

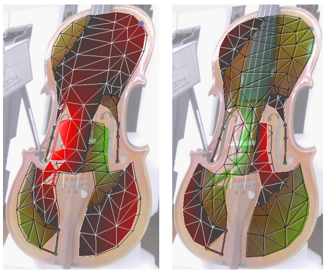


Figure 19. Helmholtz resonance mode of the top plate of the tested violin without any soundpost around 237-240 Hz.

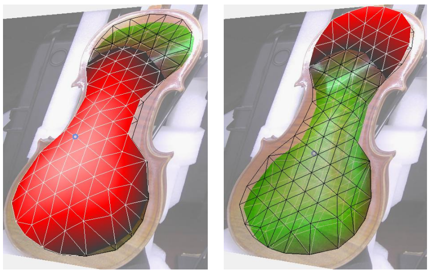


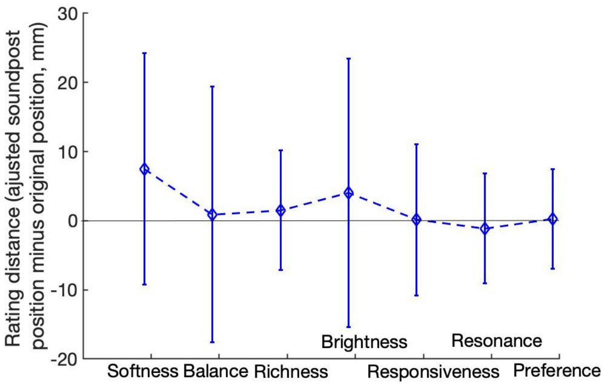
Figure 20. Helmholtz resonance mode of the top plate of the tested violin without any soundpost around 237-240 Hz.

4.2 Results of Perception Experiments

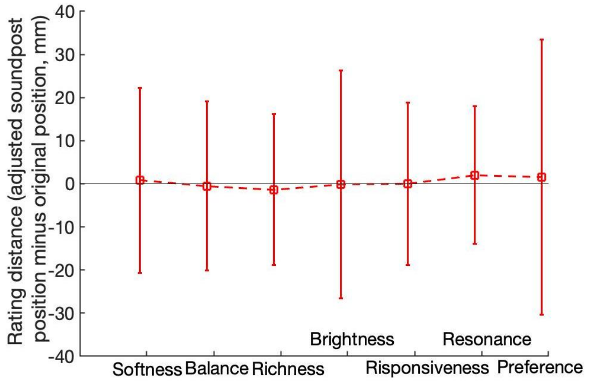
As stated in the experiment design, subjects rated the violin in each pair of trials of one phase on the same scales, thus the rating distance between the two trials meant the perceptual difference of the violin with the two soundpost positions. From Table 2, we can see that there is always an original position in any pair of trials. The ratings were measured by a ruler with the unit of millimeter. For any criterion, the rating distance was calculated in the way that the rating of any of the shifted soundpost position minus the original soundpost position. Thus, positive distance implies higher rating of the adjusted soundpost position comparing to the original position, and negative distance means lower rating of the adjusted position for any of the criteria.

The mean value of the rating distance between the original soundpost position and any of the adjusted positions across all subjects for each criterion was calculated. The mean values with 95% confidence intervals (CIs) are shown in Figure 21. From the diagrams we can see that, when moving the soundpost upwards for 5 mm, the violin becomes softer, richer and brighter, the ratings of other criteria stay about the same comparing with the original soundpost position; when moving the soundpost downwards for 5 mm, the violin almost doesn't change for any criterion; when moving the soundpost left for 5 mm, the violin becomes less balanced, less richer, less responsive but brighter; while shifting the soundpost right for 5 mm, the violin turns softer, more balanced, more responsive and more resonant.

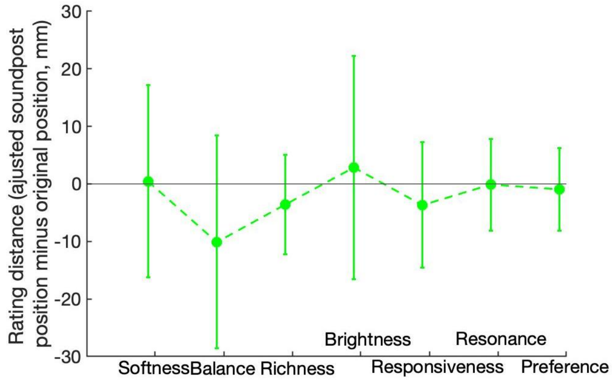
Statistical analysis was performed to test whether the mean values are significantly greater or smaller than 0, i.e., whether the variation is significant after adjusting the soundpost position. Shapiro-Wilk tests were first conducted to measure the distributions of each criterion rating distances for all subjects, and the results shows that other rating distances are normally distributed except the rating distances of the softness, brightness and responsiveness between the soundpost shifted upwards and the original position, and the rating distances of the responsiveness and resonance between the soundpost moved downwards and the original position. Based on the distribution results, one-sample t-tests and one-sample Wilcoxon signed rank tests were carried out on the normal distributed rating distances and non-normal distributed rating distances respectively. The results indicates all rating distances are not significant except the softness rating distances between the soundpost position moved upwards for 5 mm and the original position, as well as the balance rating distances between the soundpost position shifted right for 5 mm and the original position.



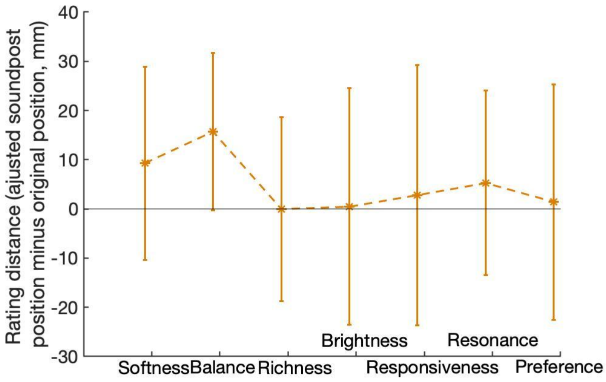
(a) Upward



(b) Downward



(c) Left



(d) Right

Figure 21. Average rating distance between the original soundpost position and any of the adjusted position across all subjects for each criterion. (Error bar corresponds to 95% CIs of the mean).

5. CONCLUSIONS

In this study, we investigated the influences of the soundpost position adjustment on the violin through both acoustic measurements and perceptual experiment. A new-designed adjustable soundpost was employed for the experiments.

In the acoustic measurements, we measured seven different soundpost conditions: wooden soundpost in the original position; adjustable soundpost in the original position; adjustable soundpost in four shifted positions: 5 mm left, right, upwards and downwards from the original position; no soundpost. Due to limited space, only one important vibration mode——the Helmholtz mode of the top and back plates is reported. In the original position, the vibrational behaviour of the violin with the wooden soundpost is quite similar to that with the adjustable soundpost around Helmholtz resonance. The air mode of the violin changes within a very narrow range while shifting the adjustable soundpost downward. But the alternations of plates' vibration are very obvious while moving the adjustable soundpost left or right. The noticeable changes in the back plate also occur while shifting the adjustable soundpost upward. By removing the soundpost, the vibration modes of the violin show up more symmetric characteristics.

In the perceptual experiments, eight subjects compared the violin with soundpost in the original position and each of the four adjusted position (corresponded to the four adjusted position in the acoustic measurements) by playing evaluation. The results shows that the violin becomes significantly softer when moving the soundpost upwards for 5 mm, and more balanced while shifting the soundpost right for 5 mm comparing with the original soundpost position. All the other variations are not significant however. The results contradicts Jansson's opinions [2]. He stated that the timbre of the violin turns softer when moving the soundpost downwards for 5mm, and the violin sound becomes lighter while placing the soundpost right for 5mm in previous studies.

Acknowledgments

Our sincere thanks are due to violin maker Fangyu Shi and all subjects for their contribution to this study. We also would like to thank Claudia Fritz for helpful discussions. This study was supported by the National Natural Science Foundation of China (NSFC) Grant.

6. REFERENCES

[1] E.V. Jansson, N-E. Molin, and H. Sundin, "Resonance of a Violin Body Studied by Hologram Interferometry and Acoustical Methods," in *Physica Scripta*, vol. 2, pp. 243-256, 1970.

[2] E.V. Jansson, *Acoustics for Violin and Guitar Makers*, 2002. [Online]. Available: http://www.speech.kth.se/music/acviguit4/part7.pdf [Accessed: Dec. 20, 2020].

[3] H. O. Saldner, N-E. Molin, and Jansson, E.V. "Vibration Modes of the Violin Forced via the Bridge and Action of the Soundpost", *Journal of the Acoustics Society of America*, 100 (2), pp. 1168-1177, 1996.

[4] G. Bissinger, "Some Mechanical and Acoustical Consequences of the Violin Soundpost", *Journal of the Acoustics Society of America*, 97 (5), pp. 3154-3164, 1994.

[5] M. C. Nadarajah, "The mechanics of the soundpost in the violin," Doctoral dissertation, University of Cambridge, Cambridge, United Kingdom, 2018.

[6] C. Gough, "A Violin Shell Model: Vibrational Modes and Acoustics", *Journal of the Acoustics Society of America*, 137 (3), pp. 1210-1225, 2015.

[7] L. Fu, C. Fritz, and G. Scavone, "Perception of violin soundpost tightness through playing and listening tests", *Journal of the Acoustical Society of America*, 150(1), pp:540-550, 2021.

[8] C. Saitis, B. L. Giordano, C. Fritz, and G. P. Scavone, "Perceptual evaluation of violins: A quantitative analysis of preference judgments by experienced players," *Journal of the Acoustical Society of America,* 132 (6), pp: 4002-4012. 2012.

BOWED STRING TRANSIENTS: ENHANCED MODELLING OF FRICTION AND FINITE-WIDTH BOWS

Jim Woodhouse
Cambridge University Engineering Department, Trumpington St, Cambridge CB2 1PZ, UK
jw12@cam.ac.uk

ABSTRACT

Models of bowed-string motion have been developed over many years, but accurate simulation of transients remains a challenge. Reliable simulation-based explorations of playability questions will only become possible when this challenge has been met. This paper will present recent developments in friction modelling, and incorporate these into an enhanced model of finite-width bowing. The friction modelling draws on earlier work with a temperature-based model, extended in response to detailed comparisons with measured waveforms. The results will be illustrated by a wide selection of Guettler diagrams: measured on a cello string with a rigid rod "bow" and a conventional bow; simulated for a violin string using point-bow models with a variety of friction models; and simulated with the same string and friction models with a finite-width bow. The results show some promise. None of the comparisons gives a perfect qualitative match to the measurements, but those measurements themselves show obvious variation between nominal repeated runs using an automated bowing machine. The real bowed string shows a significant level of sensitive dependence on initial conditions, suggesting that it may be formally on the edge of chaotic behaviour, so that there are limits to what should be expected when comparing with simulations.

1. INTRODUCTION

Theoretical models of a bowed string stretch back over 100 years, to the work of Raman. In the 1970s efficient computer simulation methods became available, and since then simulation has been used to explore issues of "playability": which steady regimes are possible under given bowing conditions, and also which of those regimes can be accessed with a given transient bowing gesture. In principle, such studies could shed light on what a string player may mean when they describe an instrument as hard or easy to play.

However, for such a study to be useful, the simulation model has to capture transient details reliably. This has proved to be a challenging task. There are many ingredients of the physical system that play a role: the transverse

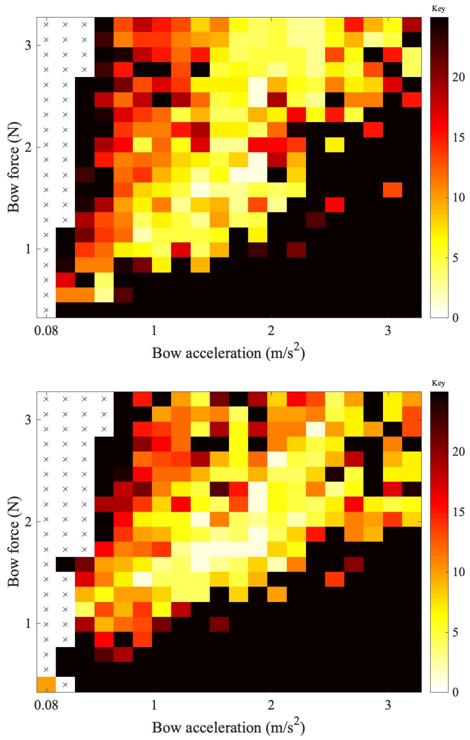


Figure 1. Guettler diagrams measured on the open D string of a cello, using a rosin-coated rod (upper plot) and a normal bow (lower plot). The bow position was $\beta = 0.0899$. Colours indicate length of transient in nominal period-lengths.

and torsional vibration of the string, coupling to the instrument body at the ends of the string, the dynamics of the bow, the finite width of bow-hair in contact with the string, and, most challenging of all, a constitutive law governing the friction force between bow and string [1]. This paper will present newly extended modelling of friction and of a finite-width bow, and compare the resulting simulations with detailed measurements of bowed-string behaviour.

For a review of the earlier history of bowed-string research, see for example [2]. Further detail on the material of the present paper can be found in Chapter 9 of https://euphonics.org

2. MEASURED GUETTLER DIAGRAMS

The specific measurements to be used are the result of using a computer-controlled bowing machine to generate

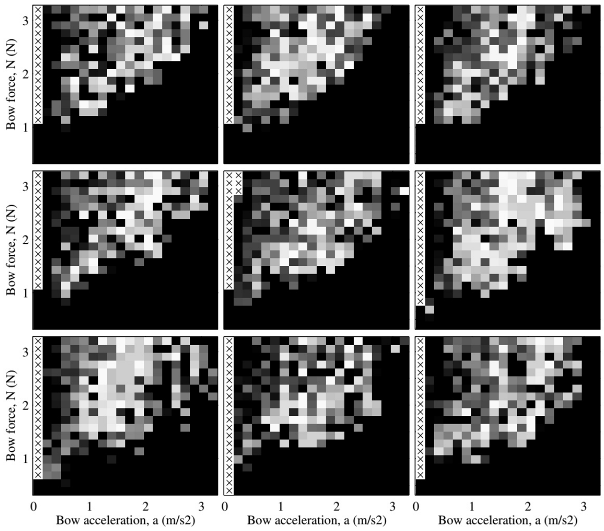


Figure 2. Nine repeated measurements of a particular Guettler diagram, demonstrating sensitivity of details. These measurements used a rosin-coated rod.

a "Guettler diagram". The Guettler diagram shows behaviour following a particular family of bowing gestures, in which the bow is first placed in contact with the string with a specified normal force, then it moves from rest with a given constant acceleration. Guettler showed that there is a wedge-shaped region in the force-acceleration plane within which it may be possible to execute a "perfect transient" in which Helmholtz motion is established immediately [3].

Measurements were performed by Galluzzo using a bowing machine to generate a 20×20 grid of transients in the Guettler plane [4]. Two typical sets of his results are shown in Fig. 1: one using a normal bow, the other using a rosin-coated rod to give "point contact". Coloured pixels indicate transients that gave rise to Helmholtz motion, with the colour indicating the transient length. Black pixels were unsuccessful, white pixels marked with crosses had insufficient data for the post-processing algorithm to operate. A hint of Guettler's wedge can be seen, but the image is very "speckly", indicating sensitive variation.

Furthermore, repeated measurements of the same Guettler plane gave different details in this speckly terrain: see Fig. 2 for an example (using the rosin-coated rod). It seems likely that the bowed string is on the edge of chaotic behaviour: not in the colloquial sense of completely uncontrollable motion, but in the mathematical sense of showing evidence of sensitive dependence on initial conditions, a classic hallmark of chaos. This sensitivity may be implicated in the tendency of a bowed string to show cycle-by-cycle variability ("jitter"), with possible perceptual consequences (for some background on this question, see [5]).

3. FRICTION MODELS

3.1 The friction-curve model

The earliest models of a bowed string assumed that friction force was governed by a simple "law" that we will call the friction-curve model. During sticking, the friction force can take any value between the limits of static friction. During sliding, the friction force is assumed to be a nonlinear function of the instantaneous sliding speed, typically with a functional form fitted to measurements of friction during steady sliding, as shown in Fig. 3.

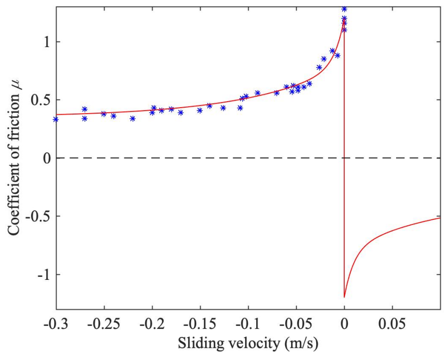


Figure 3. The friction curve: measurements by Smith [6] of steady-sliding friction of rosin (stars) and a curve-fit (red line).

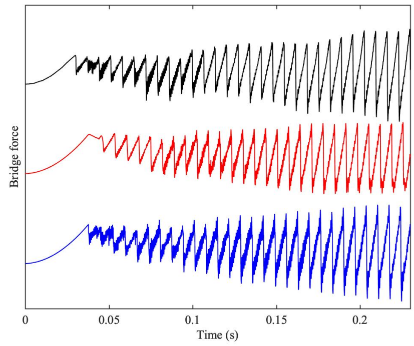


Figure 4. Matching Guettler transients with a single-point bow: measured with a rosin-coated rod (black), simulated with the thermal friction model (red) and simulated with the modified thermal friction model (blue).

3.2 The thermal model

Simulations using the friction-curve model were found to give poor results for transients, so an improved model was developed based on the temperature-dependent behaviour of violin rosin [7]. The thermal model assumes a temperature-dependent yield strength of the rosin layer, and the particular form of this dependence was deduced by fitting to the steady-sliding behaviour shown in Fig. 3. Details, including parameter values, can be found in [7].

3.3 The modified thermal model

Figure 4 shows the bridge-force waveform of a simulated transient of the cello D string with the thermal model (red), compared with the corresponding measurement (black). Aspects of the behaviour match quite well, but there is one feature that this model cannot reproduce. Both waveforms show the string initially sticking, leading to a parabolic

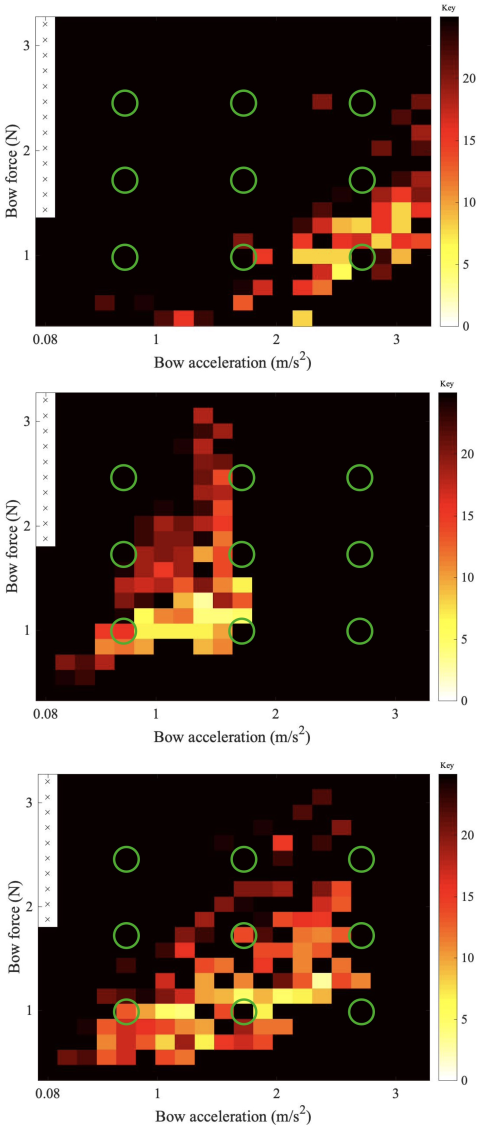


Figure 5. Guettler diagrams for a violin G string simulated using a point bow with three friction models. Top: friction-curve model; middle: thermal model; bottom: modified thermal friction model. The green circles mark waveforms illustrated in Fig. 9. The bow position is $\beta = 0.0899$.

bridge force as the bow accelerates. The first slip occurs when the parabolic curve ends. In the measurement, the bridge force jumps at first slip. But the thermal model cannot ever predict such a sudden change in bridge force: the rosin in the contact region cannot start to heat up until slip occurs, so the initial slipping gives smooth variation of bridge force while that heating occurs.

To allow a jump, a modified thermal model is proposed, which combines elements of both earlier models. Friction force is determined by both sliding speed and temperature, the variation with sliding speed being assumed to be a more gentle version of Fig. 3 with a significantly smaller difference between peak sticking friction and asymptotic high-speed sliding friction. The temperature-dependent yield stress for the thermal part of the model is recalculated so that the model still matches the steady-sliding measurements. A version of this model gives the waveform plotted in blue in Fig. 4: parameter values for the velocity-dependent function have been chosen so that the waveform shows a plausible initial jump at first slip. Details of the model and the chosen parameter values can be found in Section 9.6.2 of `https://euphonics.org`.

Guettler diagrams generated from grids of simulated transients are shown for all three friction models in Fig. 5. For reasons to be explained in the next section, these runs have used parameter values appropriate to a violin G string rather than the cello string used by Galluzzo for his measurements. In all three cases, a single point of contact between "bow" and string is assumed. It is immediately clear that the three models give very different predicted patterns. However, note that they cannot be compared directly to the measurement in Fig. 1, which was for a cello string rather than a violin string.

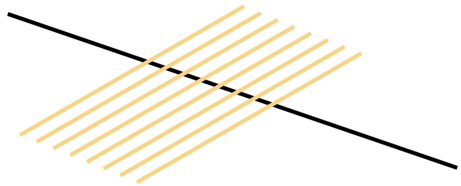


Figure 6. Schematic sketch of a set of discrete "bow-hairs", regularly spaced across the chosen section of the string.

4. FINITE-WIDTH BOWING

Another important aspect of the physics of normal bowing is the finite width of bow-hair in contact with the string. This problem was modelled by Pitteroff [8], but that work was done at a time when the rest of the bowed-string model was quite primitive. That modelling is now brought up to date, and in particular it is modified to allow all three candidate friction models to be used.

The principle of the model is simple, and follows Pitteroff's original work closely. As sketched in Fig. 6, a finite number of discrete "bow hairs" are uniformly distributed over the physical width of the bow-hair ribbon. These "hairs" are not assumed to be rigid: instead, measured behaviour of individual strands of horsehair is used to deduce an approximate model consisting of a parallel spring-dashpot combination. The model then uses a finite-difference approach to the section of string lying under the bow, combined with the digital waveguide method to represent the two sections of string outside the width of the bow. At each time step, outgoing waves from the two edges of the bow are converted into incoming waves by convolution with reflection functions as usual. Both transverse and torsional waves are included.

The short section of string under the bow is treated in the simplest possible way: transverse and torsional waves are each assumed to obey the simple wave equation, with the appropriate wave speeds. Local effects of damping and bending stiffness within this short length are ignored: but for the string as a whole they are allowed for via the reflection functions. Each wave equation can be represented approximately by a central-difference form for the spatial second derivative, involving the displacements of the "hairs" on either side of the one being considered, and by a backward-difference form for the time derivative. The spring-dashpot model for the "hairs" is also expressed in a finite-difference form. Details of all this can be found in reference [8]. Putting everything together, the new value of displacement of string and bow-hair at each discrete "hair" can be calculated, from a knowledge of the displacement of it and its neighbours at previous time steps together with the friction force acting on that particular "hair". We can use any of the three friction models at this stage.

The finite-difference formulation within the bow width imposes severe numerical constraints in order to ensure convergence. For this reason, it is much easier to simulate a violin string rather than a much longer cello string. This is the reason that Guettler diagrams relevant to a violin string were shown in the previous section: direct comparisons between these and predictions of the finite-width model will be presented shortly.

An example of a simulation using this approach, assuming the modified thermal friction model, is shown in Fig. 7. This is a case which produced Helmholtz motion, and the top plot shows a short extract of the resulting bridge force waveform. The middle plot shows the pattern of sticking (black) and slipping (white) across the width of the bow, as a function of time, for the same time segment. It shows the phenomenon of "differential slipping": the "hairs" closest to the bridge (top of the plot) frequently slip, while the "hairs" on the opposite side of the bow (bottom of the plot) continue to stick. Only the main "Helmholtz" slips reach all the way through the ribbon of bow-hair.

The bottom plot in this figure shows a corresponding map of the computed temperature across the width of the bow. Perhaps surprisingly, the local temperature is predicted to vary by almost 30°C within each cycle. This is a comparable range to the results shown earlier [7] for the original thermal model with a point bow.

Figure 8 shows three Guettler diagrams using the finite-width model and the three candidate friction models. The format is the same as in Fig. 5. Comparing the two is complicated, but we can summarise. For the friction-curve model, everything gets worse compared to the point-bow model. There are hardly any non-black pixels. The original thermal model, though, behaves impressively well. There is an almost solid wedge of coloured pixels, including some bright colours indicating short transients. The modified thermal model this time seems to make things worse. Although it gives a "Guettler wedge" in roughly the same place as the original thermal model, the colours are less bright and there are more black pixels mixed in.

Finally, Fig. 9 shows some of the detailed waveforms lying behind the post-processed Guettler plots. Each Guettler diagram has been annotated with a grid of green circles to mark the pixels corresponding to acceleration values 5, 11 and 17 (counting from the left-hand side), and force values 5, 10 and 15 (counting from the bottom). The waveforms for these nine cases are shown in Fig. 9, for all

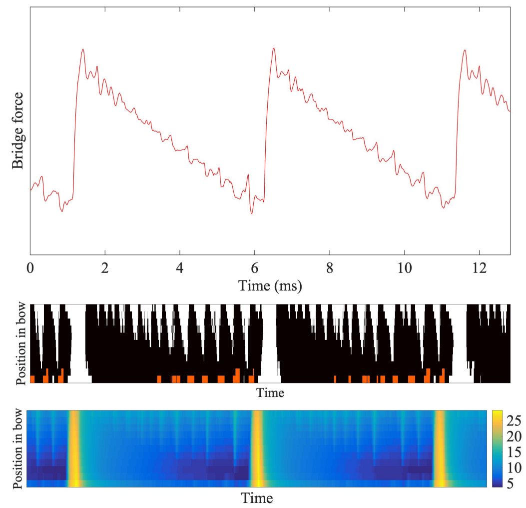


Figure 7. Simulation of Helmholtz motion with a finite-width bow using the modified thermal model. Top: bridge force waveform; middle: map showing the distribution of sticking and slipping across the width of the bow, as a function of time; bottom: corresponding map of the temperature distribution across the bow, with the brightest yellow denoting 30°C above ambient.

six models. The nine panels of that figure are laid out in the same arrangement as the grid of green circles. In each panel, results for the six different models are identified by plot colour, as described in the caption.

5. CONCLUSIONS

The picture resulting from all these simulations is tantalising. The results reinforce the message that the friction-curve model gives very poor transient performance, and is not good enough for any serious study of playability issues. The thermal models perform much better, and it looks as if the finite-width model may have eliminated a problem with the point-bow model that resulted in solidly black pixels in the right-hand half of the middle plot of Fig. 5. The differences between the two versions of the thermal model are relatively slight, so even though neither model is likely to be correct in full physical detail, we are perhaps converging on a model that is fairly robust in the face of changing model details. This gives encouragement for future work using these models to study playability.

Unfortunately, the Guettler diagrams shown in Figs. 5 and 8 cannot be compared directly with the measured versions in Figs. 1 because the measurements were on a cello string, but for computational reasons the finite-width simulations used here apply to a violin G string. A direct comparison is a task for future research: it will require either measurements on a violin string, or further refinement of the computational model to allow converged simulations of a cello string.

Acknowledgments

The development of the bowing machine and measurements

of the Guettler diagram were carried out by Paul Galluzzo [4]. Extensions to bowed-string simulation by Hossein Mansour formed a valuable part of this work (see [9]).

6. REFERENCES

[1] P. M. Galluzzo, J. Woodhouse, and H. Mansour, "Assessing friction laws for simulating bowed-string motion," *Acta Acustica united with Acustica*, vol. 103, no. 1, pp. 1080–1099, 2017.

[2] J. Woodhouse and P. M. Galluzzo, "The bowed string as we know it today," *Acta Acustica united with Acustica*, vol. 90, no. 4, pp. 579–589, 2004.

[3] K. Guettler, "On the creation of the Helmholtz motion in bowed strings," *Acta Acustica united with Acustica*, vol. 88, no. 6, pp. 970–985, 2002.

[4] P. M. Galluzzo and J. Woodhouse, "High-performance bowing machine tests of bowed-string transients," *Acta Acustica united with Acustica*, vol. 100, no. 1, pp. 139–153, 2014.

[5] M. E. Mcintyre, R. T. Schumacher, and J. Woodhouse, "Aperiodicity in bowed-string motion," *Acustica*, vol. 49, no. 1, pp. 13–32, 1981.

[6] J. H. Smith and J. Woodhouse, "The tribology of rosin," *Journal of the Mechanics and Physics of Solids*, vol. 48, no. 8, pp. 1633–1681, 2000.

[7] J. Woodhouse, "Bowed string simulation using a thermal friction model," *Acta Acustica united with Acustica*, vol. 89, no. 2, pp. 355–368, 2003.

[8] R. Pitteroff and J. Woodhouse, "Mechanics of the contact area between a violin bow and a string. Part II: Simulating the bowed string," *Acta Acustica united with Acustica*, vol. 84, no. 3, pp. 744–757, 1998.

[9] H. Mansour, J. Woodhouse, and G. Scavone, "Enhanced wave-based modelling of musical strings. Part 2: Bowed strings," *Acta Acustica united with Acustica*, vol. 102, no. 6, pp. 1094–1107, 2016.

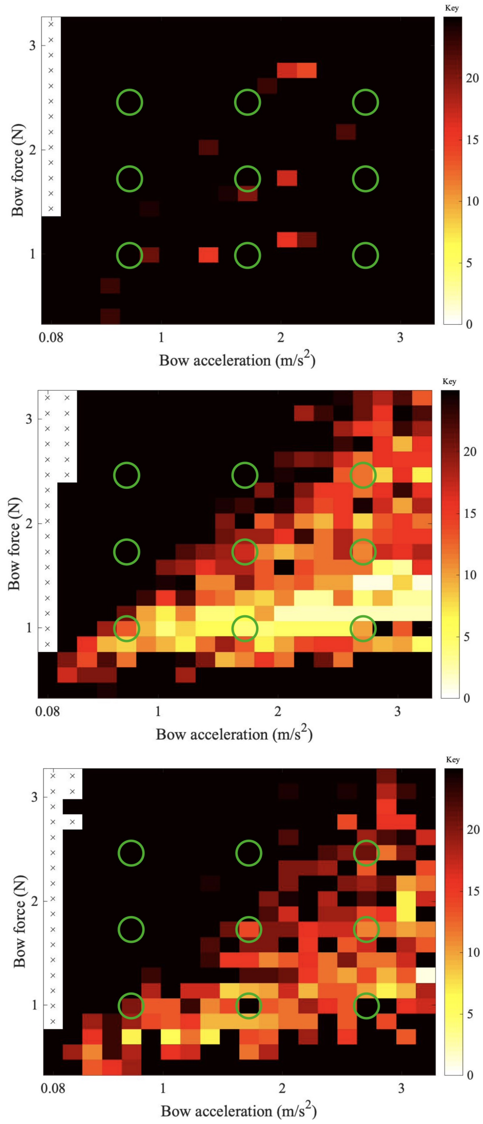


Figure 8. Guettler diagrams for a violin G string simulated using a finite-width bow with three friction models. Top: friction-curve model; middle: thermal model; bottom: modified thermal friction model. The green circles mark waveforms illustrated in Fig. 9. The bow position is $\beta = 0.0899$ as in Fig. 5.

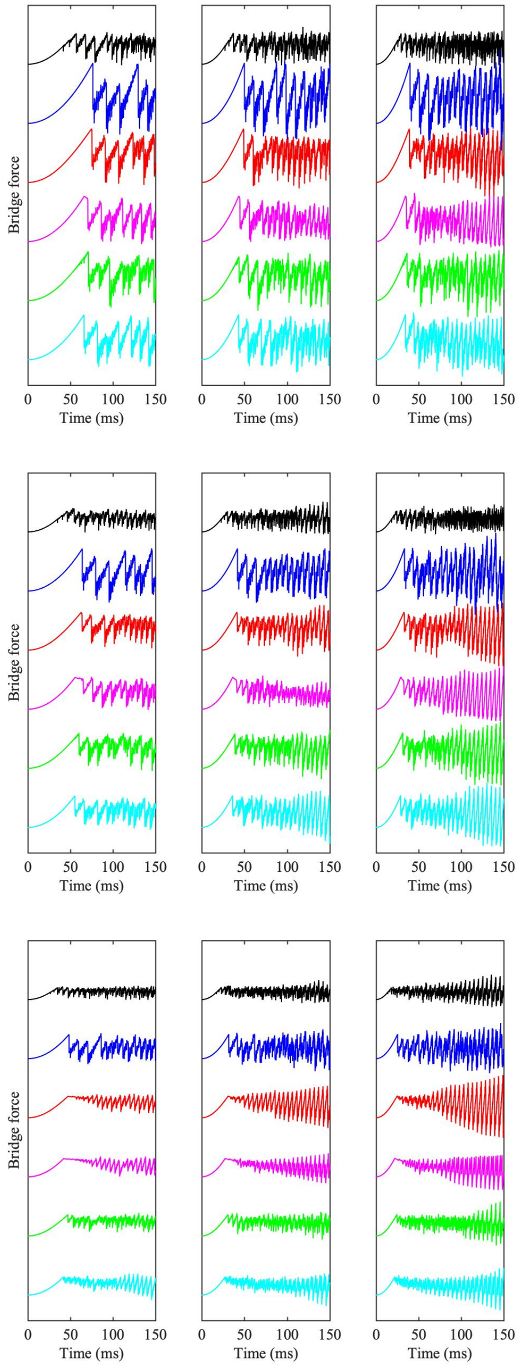


Figure 9. Examples of Guettler transients corresponding to the points marked in green in Guettler diagrams shown earlier. Each box contains matching transients from the six models. All the black curves are for the finite-width model using the friction curve, laid out geographically in the same grid as the green circles in Fig. 8. The blue curves show results for the point-bow model using the friction curve, the top plot in Fig. 5. The red curves show results for the finite-width, thermal model (Fig. 8 middle). The magenta curves show results for the point-bow, thermal model (Fig. 5 middle). The green curves show results for the finite-width, modified thermal model (Fig. 8 bottom). The cyan curves show results for the point-bow, modified thermal model (Fig. 5 bottom).

Helical Spring Mode Grouping

Jacob McQuillan and Maarten van Walstijn
SARC, Queen's University Belfast, UK
jmcquillan15@qub.ac.uk

ABSTRACT

The main vibrational elements of a spring reverb tank are helical springs. The coupling between bending, longitudinal, and torsional oscillations in a single spring gives rise to complex vibrational behaviour characterised by several distinct echo patterns in measured impulse responses. For modelling purposes, a reduced representation that covers only the hearing range can be obtained by casting the equations in modal form. Previous attempts to explain the various features of measured impulse responses have largely focused on dispersion relation analysis. However, addressing open questions on driving, pick-up, and damping mechanisms requires also considering modal amplitudes. This work applies a pseudospectral method to a thin spring version of Wittrick's twelve equations, framing an eigenvalue problem from which modal parameters are extracted. To investigate the influence of model parameters on the system response, we propose grouping modes across the wave number axis, which requires extracting wave numbers from mode shapes. This allows separate visualisation and analysis of the various echo patterns in the modelled impulse response. Preliminary results give new perspectives on how transition frequencies arise, and on the significance of specific mode groups.

1. INTRODUCTION

Spring reverb tanks are electromechanical units originally designed for Hammond organs as a compact way to incorporate reverberation [1]. Tanks typically consist of one or several helical springs connected in series or parallel with an electromagnetic transducer at each end forming reciprocal input and output mechanisms. This includes a U-shaped metal core passing through a coil of wire connected to the electric input signal and within its air gap sits a small magnetic bead connected directly to the spring. For a detailed view of the tank geometry, see [2, 3].

The complex nature of helical spring vibrations, in particular the coupling between polarisations, yields a dispersive effect with distinct aural differences to room reverberation. A typical measured spring response is visualised in Fig. 1 and can be split into two regions according to a transition frequency dependent on the spring geometry. The lower region includes a series of dispersive echoes (or *chirps*) that progressively blur and in the upper region a more regular set of echoes with a higher repetition rate is visible. Various analyses of measured responses are yet to draw decisive conclusions on certain components of spectrograms, e.g., sets of faint secondary chirps are evident and the main chirps are smeared at low frequencies [4–6].

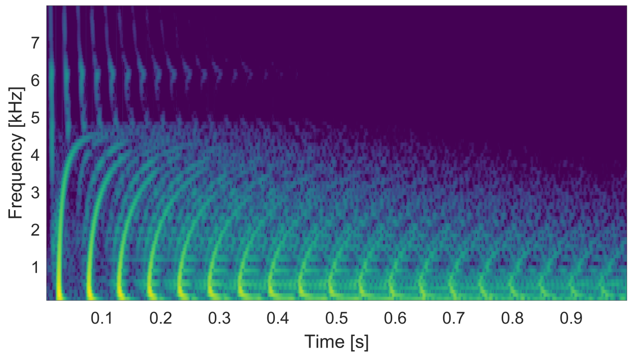


Figure 1. *Impulse response of a single spring in an Accutronics 2EB2C1B reverb tank.*

In terms of modelling, one of two sets of equations are typically the starting point in this context: a simplified helical spring model neglecting helix angle [7], or a more complex model from Wittrick derived as an extension to curvature of the Timoshenko beam [8]. For the latter, removal of coupling terms not affecting the hearing range for typical reverb tank spring geometries allows derivation of a two-variable *thin spring model* [6, 9].

In past research, finite difference time domain (FDTD) methods are commonly employed for both the simple helical spring model [5, 10] and the more complex model [6, 9], although magnetic beads have only been modelled at the input and output for the former [10]. The FDTD approaches capture the overall features of measurements well, though the required amount of computation for achieving accuracy over the hearing range is relatively high [9, 11].

An alternative approach employs higher order finite difference (HOFD) schemes paired with a large-matrix diagonalisation to arrive at a modal formulation, with applications both to the simple model [11] and the more complex model [2] (where the finite difference scheme for the latter is semi-discrete). Dispersion relation analysis indicates that the modal structure captures the vibrational behaviour of a helical spring across the hearing range, however, modelled impulse responses show that mode amplitudes are not representative of measurements. As such, attaining a

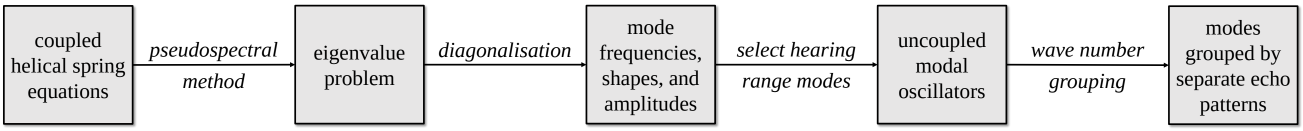


Figure 2. *Overview of the path from helical spring equations to a modal structure.*

closer match to measurements requires incorporating magnetic beads coupled to either end of the helical spring, as was previously done for the simple model [10]. However, applying the HOFD approach to such a coupled system is uncertain as ghost nodes arising from the high order centred approximations were addressed only by application of simple boundary conditions.

As a preliminary step towards the numerical simulation of a model including beads coupled to the spring, this paper proposes a pseudospectral method to arrive at a modal formulation of the thin spring model. Such an approach will more readily extend to models with several connected elements as it avoids the ghost nodes arising in HOFD methods.

Furthermore, analysis of the envisaged model with beads would not be fully characterised by its dispersion relation since this does not account for the input and output. As such, we propose grouping modes by wave number to facilitate future investigations of open questions around modelling choices (e.g., geometry of the bead and spring coupling). The perspective this grouping will provide for the mode amplitudes directly correlates to the perceptual significance of separate echo patterns in simulated responses for different model choices. Fig. 2 provides an overview of the proposed approach forming a semi-discrete helical spring model via a pseudospectral method then grouping modes by wave number after extracting modal parameters from the resultant eigenvalue problem. In addition to the above motivations for a modal approach, modal engines are also well suited to real-time implementations since their structure allows parallel computing to be leveraged.

The coupled equations here are a new four-variable thin spring model that for zero curvature describes bending, longitudinal, and torsional vibrations of a straight beam. This will facilitate future incorporation of the beads since they will be connected to the spring via an unwound wire section and these three elements would ideally be modelled with the same equations. However, the two-variable thin spring model does not readily reduce to Euler-Bernoulli beam equations in the absence of curvature, thus the new four-variable form is introduced here.

2. HELICAL SPRING MODEL

Wittrick's twelve equations [8] are written here in vector-matrix form:

$$\rho A \partial_t^2 \mathbf{x} = \mathbf{A}\mathbf{p} + \mathbf{B}\mathbf{q}, \tag{1}$$

$$\mathbf{C}\partial_t^2 \mathbf{y} = \mathbf{D}\mathbf{m} - \mathbf{B}\mathbf{p} + \partial_s \mathbf{q}, \tag{2}$$

$$\mathbf{A}\mathbf{x} = \mathbf{H}\boldsymbol{\theta} - \mathbf{B}\mathbf{y} + \frac{\mathbf{p}}{GA\gamma}, \tag{3}$$

$$\partial_s \mathbf{y} = \mathbf{B}\mathbf{x} - \mathbf{D}\boldsymbol{\theta} + \mathbf{J}\mathbf{q}, \tag{4}$$

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{D}^{\mathrm{T}}\mathbf{y} + \frac{\mathbf{m}}{EI}, \tag{5}$$

$$\mathbf{A}\mathbf{m} = \mathbf{H}\mathbf{p} + \mathbf{D}^{\mathrm{T}}\mathbf{q} + \rho I \partial_t^2 \boldsymbol{\theta}, \tag{6}$$

describing vibrations of a helical spring of circular cross-section, where

$$\mathbf{x} = \begin{bmatrix} u \\ v \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} w \\ \theta_w \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} p_u \\ p_v \end{bmatrix}, \tag{7}$$

$$\mathbf{m} = \begin{bmatrix} m_u \\ m_v \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \theta_u \\ \theta_v \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} p_w \\ m_w \end{bmatrix}, \tag{8}$$

$$\mathbf{A} = \begin{bmatrix} \partial_s & -\tau \\ \tau & \partial_s \end{bmatrix}, \; \mathbf{B} = \begin{bmatrix} \kappa & 0 \\ 0 & 0 \end{bmatrix}, \; \mathbf{C} = \begin{bmatrix} \rho A & 0 \\ 0 & 2\rho I \end{bmatrix}, \tag{9}$$

$$\mathbf{D} = \begin{bmatrix} 0 & 0 \\ -\kappa & 0 \end{bmatrix}, \; \mathbf{H} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \; \mathbf{J} = \begin{bmatrix} \frac{1}{EA} & 0 \\ 0 & \frac{1}{2GI} \end{bmatrix}. \tag{10}$$

The system is defined for displacements u, v, and w, rotation angles $\theta_{u,v,w}$, forces $p_{u,v,w}$, and moments $m_{u,v,w}$ for three orthogonal polarisations: two transverse (u and v) and one longitudinal (w). All twelve variables are functions of the wire-axis coordinate s and time t: ∂_s^n and ∂_t^n represent n^{th} order partial derivative operators for space and time, respectively, and the spatial coordinate s exists on the domain $[0, L]$ for an unwound length L (m). The physical parameters are: the wire radius r (m), the coil radius R (m), the Young's modulus E (Pa), the shear modulus G (Pa), the cross-sectional area $A = \pi r^2$ (m^2), the shear area correction γ, the transverse moments of inertia $I = \pi r^4/4$ (m^4), and the material density ρ ($\mathrm{kg/m}^3$). Helix curvature and tortuosity are defined as $\kappa = \cos^2(\alpha)/R$ and $\tau = \kappa \tan(\alpha)$, respectively, where α is the spring helix angle.

The above system models helical spring vibrations well past the hearing range and for typical reverb tank geometries removing $\mathbf{p}/GA\gamma$, $\rho I \partial_t^2 \boldsymbol{\theta}$, $2\rho I \partial_t^2 \theta_w$, and p_w/EA is shown not to affect the hearing range behaviour [9]. Various substitutions then yield the two-variable thin spring model [6, 9].

Here, motivations are informed by long term aims of modelling the input-output components as stepped beams coupled to either end of the spring. As such, of the negligible coupling terms removed above, the longitudinal components are now retained, and concatenating $\mathbf{x}$ and $\mathbf{y}$ allows elimination of $\mathbf{p}$, $\mathbf{m}$, $\boldsymbol{\theta}$, and $\mathbf{q}$ to yield a four-variable model:

$$\partial_t^2 \mathbf{z} = \mathbf{W}\mathbf{z}, \tag{11}$$

where

$$\mathbf{z} = \begin{bmatrix} u & v & w & \theta_w \end{bmatrix}^{\mathrm{T}}, \quad \mathbf{W} = \begin{bmatrix} \mathbf{S} & \mathbf{T} \\ \mathbf{U} & \mathbf{V} \end{bmatrix}, \tag{12}$$

and

$$\mathbf{S} = \frac{1}{\rho A}\left(\mathbf{AQ}+\mathbf{BN}\right), \quad \mathbf{T} = \frac{1}{\rho A}\left(\mathbf{AR}+\mathbf{BP}\right), \tag{13}$$

$$\mathbf{U} = \mathbf{C}^{-1}\left(-EI\mathbf{DL}-\mathbf{BQ}+\partial_s\mathbf{N}\right), \tag{14}$$

$$\mathbf{V} = \mathbf{C}^{-1}\left(-EI\mathbf{DM}-\mathbf{BR}+\partial_s\mathbf{P}\right). \tag{15}$$

Matrices in (13–15) are built using a 2×2 identity matrix $\mathbf{I}$ and matrices in (9,10):

$$\mathbf{N}=\mathbf{J}^{-1}(-\mathbf{B}-\mathbf{DHA}), \;\; \mathbf{P}=\mathbf{J}^{-1}(\mathbf{I}\partial_s-\mathbf{DHB}), \tag{16}$$

$$\mathbf{L}=\mathbf{AHA}, \quad \mathbf{Q}=EI\mathbf{HAL}+\mathbf{HD}^{\mathrm{T}}\mathbf{N}, \tag{17}$$

$$\mathbf{M}=\mathbf{AHB}+\mathbf{D}^{\mathrm{T}}, \quad \mathbf{R}=EI\mathbf{HAM}+\mathbf{HD}^{\mathrm{T}}\mathbf{P}. \tag{18}$$

Evaluating $\mathbf{W}$ for the case of zero curvature yields an uncoupled beam:

$$\partial_t^2\begin{bmatrix} u \\ v \\ w \\ \theta_w \end{bmatrix} = \begin{bmatrix} \frac{-EI}{\rho A}\partial_s^4 & 0 & 0 & 0 \\ 0 & \frac{-EI}{\rho A}\partial_s^4 & 0 & 0 \\ 0 & 0 & \frac{E}{\rho}\partial_s^2 & 0 \\ 0 & 0 & 0 & \frac{G}{\rho}\partial_s^2 \end{bmatrix}\begin{bmatrix} u \\ v \\ w \\ \theta_w \end{bmatrix}. \tag{19}$$

2.1 Boundary Conditions

The endpoints in a reverb tank consist of a short wire held in a pinned condition which connects to the magnetic beads coupled to the helical spring. For a model omitting beads, a set of lossless boundary conditions is required. One option is:

$$\partial_s u\,(0,t)=\partial_s^3 u\,(0,t)=\partial_s u\,(L,t)=\partial_s^3 u\,(L,t)=0, \tag{20}$$

$$v\,(0,t) = \partial_s^2 v\,(0,t) = v\,(L,t) = \partial_s^2 v\,(L,t) = 0, \tag{21}$$

$$w\,(0,t) = \partial_s^2 w\,(0,t) = w\,(L,t) = \partial_s^2 w\,(L,t) = 0, \tag{22}$$

$$\theta_w\,(0,t)=\partial_s^2\theta_w\,(0,t)=\theta_w\,(L,t)=\partial_s^2\theta_w\,(L,t)=0, \tag{23}$$

which is derived using the boundary energy term defined in [9] for the thin spring model and consisting of six products composed of the original twelve variables. Setting this term to zero via the same choices as in [2] yields the above conditions when applied to the four-variable model.

2.2 Excitation and Pick-Up

Experimental investigations on a reverb tank's excitation mechanism indicate that the magnetic bead is driven primarily through twisting [3]. The linear coupling in springs dictates that all polarisations will be excited to some extent and, at the output all polarisations have an influence. Previous research indicates that simple mechanisms defining the excitation and pick-up in regions at the start and end of the spring are unlikely to be similar to the real-life system [2].

As such, the choice of which polarisation to excite/pick-up is left open here: four options are explored, examining the cases of exciting and listening to the same variable at each end. Since the beads are not modelled, the excitation is applied uniformly over a short (1 cm) region at the start of the spring and the resultant driven model is written as

$$\partial_t^2\mathbf{z} = \mathbf{Wz} + \mathbf{k}_{\mathrm{E}}F_{\mathrm{E}}, \tag{24}$$

for a given force F_{E}, where $\mathbf{k}_{\mathrm{E}} = \psi_{\mathrm{E}}\,(s)\,\boldsymbol{\chi}_{\mathrm{E}}$ using a distribution function ψ_{E}:

$$\psi_{\mathrm{E}}\,(s) = \begin{cases} \frac{1}{\eta} & : \quad 0 < s \leq \eta \\ 0 & : \quad \text{otherwise} \end{cases}, \tag{25}$$

for a given width η (in this case $\eta = 1$ cm). The vector $\boldsymbol{\chi}_{\mathrm{E}}$ defines the choice of which polarisation to excite, e.g., choosing θ_w, $\boldsymbol{\chi}_{\mathrm{E}} = \begin{bmatrix}0 & 0 & 0 & 1\end{bmatrix}^{\mathrm{T}}$.

Following the reciprocal nature of the transducers, the pick-up is obtained by "listening" to the same region at the other end of the spring:

$$\psi_{\mathrm{P}}\,(s) = \psi_{\mathrm{E}}\,(L-s)\,, \tag{26}$$

and the pick-up force is defined as

$$F_{\mathrm{P}} = \int_0^L \mathbf{k}_{\mathrm{P}}^{\mathrm{T}}\partial_t^2\mathbf{z}\;ds = \int_0^L \mathbf{k}_{\mathrm{P}}^{\mathrm{T}}\mathbf{Wz}\;ds, \tag{27}$$

where $\mathbf{k}_{\mathrm{P}}$ is formed using $\psi_{\mathrm{P}}\,(s)$ and $\boldsymbol{\chi}_{\mathrm{P}}$ for any of the four variables analogous to $\mathbf{k}_{\mathrm{E}}$.

2.3 Dispersion Relation

The dispersion relation of a system links temporal frequency ω to wave number β (i.e., spatial frequency) and an example is shown for a helical spring with model parameters from [5] in Fig. 3. The curves are derived exploring the ansatz $e^{\mathrm{j}(\omega t+\beta s)}$:

$$u = Ue^{\mathrm{j}(\omega t+\beta s)}, \quad v = Ve^{\mathrm{j}(\omega t+\beta s)}, \tag{28}$$

$$w = We^{\mathrm{j}(\omega t+\beta s)}, \;\; \theta_w = \Theta_w e^{\mathrm{j}(\omega t+\beta s)}, \tag{29}$$

where $\mathrm{j} = \sqrt{-1}$ and U, V, W, and Θ_w are complex amplitudes. Inserting these expressions into (11) leads to the transformation of derivative operators as

$$\partial_s \to \mathrm{j}\beta, \quad \partial_s^2 \to -\beta^2, \quad \partial_s^3 \to -\mathrm{j}\beta^3, \quad \partial_s^4 \to \beta^4. \tag{30}$$

Non-trivial solutions then occur by setting the determinant of $\mathbf{W}$ after the above transformation to zero, yielding an eighth-order equation in ω that can be evaluated for a given β.

Fig. 3(a) shows two curves in the hearing range each giving rise to two transition frequencies, where a corresponding wave number also exists. The peak of the main hump gives $(f_{\mathrm{T1}}^-, \beta_{\mathrm{T1}}^-)$ and $(f_{\mathrm{T1}}^+, \beta_{\mathrm{T1}}^+)$ for each curve, and the dip gives $(f_{\mathrm{T2}}^-, \beta_{\mathrm{T2}}^-)$ and $(f_{\mathrm{T2}}^+, \beta_{\mathrm{T2}}^+)$, where $f_{\mathrm{T2}}^- = 0$ Hz. Fig. 3(b) shows two further curves well above the hearing range. For the case of a beam ($\kappa = 0$) these curves exist in this region, albeit over a very small wave number range.

3. CHEBYSHEV PSEUDOSPECTRAL METHOD

Pseudospectral methods are explored here both for their strong convergence properties [12] and the global nature of approximations being suited to simulating systems with multiple connected components. They are a subset of spectral methods and notable books introducing the topic include [12–14]. The fundamental concept (as, e.g., in [13]) is to assume the solution to a differential equation can be

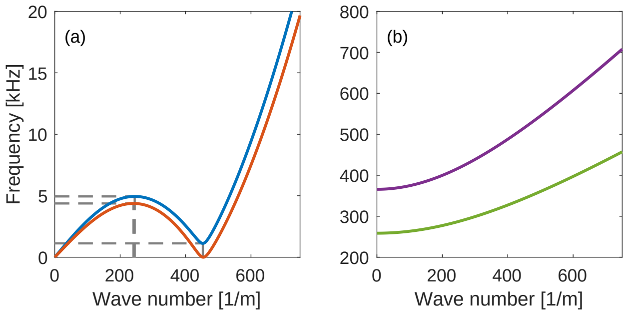


Figure 3. *Dispersion relation of the four-variable model which shows (a) two curves in the hearing range and (b) two curves well above this region. Grey lines show transition frequencies and wave numbers.*

approximated by $M+1$ basis functions, described here for an arbitrary variable:

$$\sigma(s) \approx \sigma_M(s) = \sum_{n=0}^{M} a_n T_n(s), \tag{31}$$

where a_n is a set of coefficients and first kind Chebyshev polynomials defined as

$$T_n(s) = \cos\left(n \cos^{-1}(s)\right), \tag{32}$$

are employed for the basis functions which are common for non-periodic spatial domains [13] and chosen for their applicability to any boundary conditions. Substituting (31) into the equation to be solved yields a residual function and the primary challenge lies in choosing a_n such that this is minimised. A pseudospectral method involves setting the residual to zero at a set of collocation points, i.e., partial differential equations are solved pointwise in space [13]. Such is a common choice both due to comparative simplicity and its wide applicability to problems.

Here, we employ Gauss-Chebyshev-Lobatto (GCL) points defined as

$$s_j^* = \cos\left(\frac{j\pi}{M}\right), \quad j = 0, 1 \dots M, \tag{33}$$

where s_j^* exists between 1 and -1. These are derived by evaluating the extrema of (32) and are a common choice as the nodes clustering at the boundaries avoids the Runge phenomenon seen for equispaced points in polynomial interpolation problems [12]. Re-mapping the domain to $[0, L]$ yields new grid points s_j:

$$s_j = \left(\frac{-s_j^* + 1}{2}\right) L. \tag{34}$$

For the given set of points, let f be a unique polynomial of degree M, where $f(s_j) = g_j$, a given grid function, and $0 \leq j \leq M$ [12]. In other words, for a grid function with $M+1$ points, there exists a unique polynomial of degree M that passes through said points, i.e., the Lagrange interpolating polynomial [15]. Now, the derivative of the grid function g_j can be defined using the polynomial: $h_j = g_j' = f'(s_j)$. In vector form, this reads:

$$\mathbf{h} = \boldsymbol{\mathcal{D}}\mathbf{g}. \tag{35}$$

Here, differentiation matrices are generated using Chebfun, an open-source Matlab software package [16,17]. The `diffmat.m` function generates an n^{th} order differentiation matrix $\boldsymbol{\mathcal{D}}^n$ for a given set of points (a GCL grid in this case) and re-maps to a specific domain, i.e., $[0, L]$. In practical applications we define $N = M + 1$ such that $1 \leq j \leq N$.

3.1 Helical Spring Model

Applying the pseudospectral method to (24) yields

$$\partial_t^2 \overline{\mathbf{z}} = \overline{\mathbf{W}}\overline{\mathbf{z}} + \overline{\mathbf{k}}_{\mathrm{E}} F_{\mathrm{E}}, \tag{36}$$

where

$$\overline{\mathbf{z}} = \begin{bmatrix} u_j & v_j & w_j & (\theta_w)_j \end{bmatrix}^{\mathrm{T}}, \quad \overline{\mathbf{k}}_{\mathrm{E}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \psi_{\mathrm{E},j} \end{bmatrix}^{\mathrm{T}}, \tag{37}$$

and $\overline{\mathbf{z}}$ and $\overline{\mathbf{k}}_{\mathrm{E}}$ are $(4N \times 1)$ vectors with $\mathbf{0}$ denoting a zero vector. The j subscript denotes sampling on the GCL grid: $u_j = u(s_j, t)$, and the matrix $\overline{\mathbf{W}}$ is obtained by replacing derivative operators with differentiation matrices, e.g., the first element is now written as

$$\overline{W}_{1,1} = \frac{1}{\rho A}\Big[EI\left(-\boldsymbol{\mathcal{D}}^4 + 6\tau^2\boldsymbol{\mathcal{D}}^2\right) - \mathbf{I}\left(\tau^4 EI + 2\kappa^2\tau^2 GI + \kappa^2 EA\right)\Big]. \tag{38}$$

The other fifteen elements of $\overline{\mathbf{W}}$ are transformed accordingly. Analogous to the continuous-domain, the output force is defined as

$$F_{\mathrm{P}} = \overline{\mathbf{k}}_{\mathrm{P}}^{\mathrm{T}}\overline{\mathbf{W}}\overline{\mathbf{z}}. \tag{39}$$

3.2 Boundary Conditions

Two approaches for addressing boundary conditions at this stage are basis recombination or boundary-bordering [13]. The first involves re-defining basis functions such that end conditions are inherently satisfied and the second removes collocation points to then explicitly enforce conditions [13]. While the first is convenient for specific cases, the latter is chosen here for its more general applicability to any set of boundary conditions.

Replacing collocation points with end conditions involves setting rows of $\overline{\mathbf{W}}$ to zero then explicitly enforcing boundary conditions [13]. Two conditions per end are necessary for each variable in $\overline{\mathbf{z}}$: the first and last node and the neighbouring point of each are removed, i.e., their corresponding matrix rows are zeroed. Using $\mathbf{d}_p^n$ to denote the p^{th} row of an n^{th} order differentiation matrix, the boundary conditions in (20–23) are re-written as

$$\mathbf{d}_1^1\mathbf{u} = \mathbf{d}_N^1\mathbf{u} = \mathbf{d}_1^3\mathbf{u} = \mathbf{d}_N^3\mathbf{u} = 0, \tag{40}$$

$$\begin{bmatrix}1 \dots 0\end{bmatrix}\mathbf{v} = \begin{bmatrix}0 \dots 1\end{bmatrix}\mathbf{v} = \mathbf{d}_1^2\mathbf{v} = \mathbf{d}_N^2\mathbf{v} = 0, \tag{41}$$

$$\begin{bmatrix}1 \dots 0\end{bmatrix}\mathbf{w} = \begin{bmatrix}0 \dots 1\end{bmatrix}\mathbf{w} = \mathbf{d}_1^2\mathbf{w} = \mathbf{d}_N^2\mathbf{w} = 0, \tag{42}$$

$$\begin{bmatrix}1 \dots 0\end{bmatrix}\boldsymbol{\theta}_w = \begin{bmatrix}0 \dots 1\end{bmatrix}\boldsymbol{\theta}_w = \mathbf{d}_1^2\boldsymbol{\theta}_w = \mathbf{d}_N^2\boldsymbol{\theta}_w = 0. \tag{43}$$

Each of the sixteen conditions replaces the appropriate row in $\overline{\mathbf{W}}$ and corresponding components in the left hand side vector $\overline{\mathbf{z}}$ in (36) are set to zero.

3.3 Eigenvalue Problem

To yield a regular eigenvalue problem, the unforced form of (36) must be rearranged such that the system can be solved for the interior nodes with boundary conditions enforced by applying Schur complementation [18]. The boundary and interior components are split for each variable as

$$\tilde{\mathbf{z}} = \begin{bmatrix} \overline{\mathbf{z}}_{\mathrm{B}} \\ \overline{\mathbf{z}}_{\mathrm{I}} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_{\mathrm{B}} & \mathbf{v}_{\mathrm{B}} & \mathbf{w}_{\mathrm{B}} & (\boldsymbol{\theta}_w)_{\mathrm{B}} & \mathbf{u}_{\mathrm{I}} & \mathbf{v}_{\mathrm{I}} & \mathbf{w}_{\mathrm{I}} & (\boldsymbol{\theta}_w)_{\mathrm{I}} \end{bmatrix}^{\mathrm{T}}, \tag{44}$$

where, e.g.,

$$\mathbf{u}_{\mathrm{B}} = \begin{bmatrix} u_1 & u_2 & u_{N-1} & u_N \end{bmatrix}, \quad \mathbf{u}_{\mathrm{I}} = \begin{bmatrix} u_3 \dots u_{N-2} \end{bmatrix}. \tag{45}$$

Rearranging $\overline{\mathbf{W}}$ such that it corresponds to $\tilde{\mathbf{z}}$ allows partitioning of the matrix:

$$\begin{bmatrix} \mathbf{0} \\ \partial_t^2 \overline{\mathbf{z}}_{\mathrm{I}} \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{W}}_{\mathrm{BB}} & \overline{\mathbf{W}}_{\mathrm{BI}} \\ \overline{\mathbf{W}}_{\mathrm{IB}} & \overline{\mathbf{W}}_{\mathrm{II}} \end{bmatrix} \begin{bmatrix} \overline{\mathbf{z}}_{\mathrm{B}} \\ \overline{\mathbf{z}}_{\mathrm{I}} \end{bmatrix}, \tag{46}$$

where dimensions of $\overline{\mathbf{W}}_{\mathrm{BB}}$, $\overline{\mathbf{W}}_{\mathrm{BI}}$, $\overline{\mathbf{W}}_{\mathrm{IB}}$, and $\overline{\mathbf{W}}_{\mathrm{II}}$ are 16×16, $16 \times (4N-16)$, $(4N-16) \times 16$, and $(4N-16) \times (4N-16)$, respectively. Note that $\mathbf{0}$ is a vector of sixteen zeroes, i.e., enforcing that all boundary conditions are set to zero. Expanding (46) yields

$$\mathbf{0} = \overline{\mathbf{W}}_{\mathrm{BB}} \overline{\mathbf{z}}_{\mathrm{B}} + \overline{\mathbf{W}}_{\mathrm{BI}} \overline{\mathbf{z}}_{\mathrm{I}}, \tag{47}$$

$$\partial_t^2 \overline{\mathbf{z}}_{\mathrm{I}} = \overline{\mathbf{W}}_{\mathrm{IB}} \overline{\mathbf{z}}_{\mathrm{B}} + \overline{\mathbf{W}}_{\mathrm{II}} \overline{\mathbf{z}}_{\mathrm{I}}, \tag{48}$$

and substituting (47) into (48) leads to a regular eigenvalue problem for the interior nodes with the boundary conditions enforced:

$$\partial_t^2 \overline{\mathbf{z}}_{\mathrm{I}} = \left(\overline{\mathbf{W}}_{\mathrm{II}} - \overline{\mathbf{W}}_{\mathrm{IB}} \overline{\mathbf{W}}_{\mathrm{BB}}^{-1} \overline{\mathbf{W}}_{\mathrm{BI}} \right) \overline{\mathbf{z}}_{\mathrm{I}} = \mathbf{Z} \overline{\mathbf{z}}_{\mathrm{I}}. \tag{49}$$

Corresponding to the definition of interior components in (44), the discrete distribution $\overline{\mathbf{k}}_{\mathrm{E,I}}$ is defined by removing nodes corresponding to boundary conditions and is re-introduced to the governing equation:

$$\partial_t^2 \overline{\mathbf{z}}_{\mathrm{I}} = \mathbf{Z} \overline{\mathbf{z}}_{\mathrm{I}} + \overline{\mathbf{k}}_{\mathrm{E,I}} F_{\mathrm{E}}. \tag{50}$$

4. MODAL FORMULATION

Diagonalisation of $\mathbf{Z}$ in (50) yields

$$\mathbf{Z} = \boldsymbol{\Lambda} \boldsymbol{\Omega} \boldsymbol{\Lambda}^{-1}, \tag{51}$$

where the diagonal elements of $\boldsymbol{\Omega}$ hold eigenvalues and corresponding columns of $\boldsymbol{\Lambda}$ hold eigenvectors. As laid out in [2, 11], substitution of (51) into (50) leads to a set of uncoupled ordinary differential equations:

$$\partial_t^2 \boldsymbol{\xi} = \boldsymbol{\Omega} \boldsymbol{\xi} + \mathbf{c}_{\mathrm{E}} F_{\mathrm{E}}, \tag{52}$$

where $\boldsymbol{\xi} = \boldsymbol{\Lambda}^{-1} \overline{\mathbf{z}}_{\mathrm{I}}$ holds modal displacements and $\mathbf{c}_{\mathrm{E}} = \boldsymbol{\Lambda}^{-1} \overline{\mathbf{k}}_{\mathrm{E,I}}$ holds input mode amplitudes.

4.1 Modal Parameters

Exploring the ansatz $e^{\mathrm{j}\omega t}$ and ignoring the forcing term in (52) allows extraction of resonance frequencies from the eigenvalues:

$$\omega_i = \sqrt{-\Omega_{i,i}}, \quad f_i = \omega_i / 2\pi, \tag{53}$$

where i is the mode index. The output in (39) is re-written as

$$F_{\mathrm{P}} = \overline{\mathbf{k}}_{\mathrm{P,I}}^{\mathrm{T}} \mathbf{Z} \overline{\mathbf{z}}_{\mathrm{I}} = \overline{\mathbf{k}}_{\mathrm{P,I}}^{\mathrm{T}} \mathbf{Z} \boldsymbol{\Lambda} \boldsymbol{\xi} = \mathbf{c}_{\mathrm{P}} \boldsymbol{\xi}, \tag{54}$$

where $\mathbf{c}_{\mathrm{P}} = \overline{\mathbf{k}}_{\mathrm{P,I}}^{\mathrm{T}} \boldsymbol{\Lambda} \boldsymbol{\Omega}$ holds the output mode amplitudes and, as for the input, $\mathbf{k}_{\mathrm{P,I}}$ is defined for the interior nodes. Input and output amplitudes can then be consolidated via element-by-element multiplication:

$$\mathbf{c} = \mathbf{c}_{\mathrm{E}} \odot \mathbf{c}_{\mathrm{P}}^{\mathrm{T}}. \tag{55}$$

4.2 Mode Ordering and Selection

Arranging mode frequencies by ascending value and ordering mode shapes and amplitudes accordingly is adequate for selecting hearing range modes. However, it does not facilitate splitting the modes into two groups belonging to the two curves in Fig. 3(a) because the helical spring frequencies are not unique functions of wave number. As such, the eigenvectors are employed for ordering.

Since the four-variable system is coupled, each eigenvector holds four versions of the same mode shape that differ by amplitude (with the exception of u due to the different boundary conditions). Counting the zero crossings of one mode shape for each eigenvector then allows the modes to be ordered by "wave-like number" $\tilde{\imath}$, i.e., the number of zero crossings plus one.

Because the hearing range is of interest, only modes between 20 Hz and 20 kHz need to be retained. This informs the choice of N, although for high wave numbers mode shapes start to exhibit aliasing in the middle of the GCL grid (where grid spacing is widest). For a given N, where this begins can be determined empirically as the point when the zero crossing counting starts to fail, and thus only the non-aliased modes within the hearing range are retained.

4.3 Mode Grouping

Each of the retained modes from the eigenvalue calculation falls on one of the two curves exemplified in Fig. 3(a). Fig. 4 shows mode shapes for spring parameters from the reverb tank in an Olson X-82 amplifier [5] for the first three modes of each group. Only the dominant mode shape in each eigenvector is shown as the other has a much lower amplitude. It is clear these form a harmonic series of sine functions, as would be expected for pinned boundary conditions applied to a one-dimensional system. As such, for this case $\tilde{\imath} = i$ and a modal wave number can be derived as $\beta_i = \tilde{\imath}\pi / L$.

The definition of β_i facilitates use of the dispersion relation as a measure of accuracy for the mode frequencies. Fig. 5 shows mode frequencies and wave numbers for a short spring with $N = 175$. Spring parameters are again

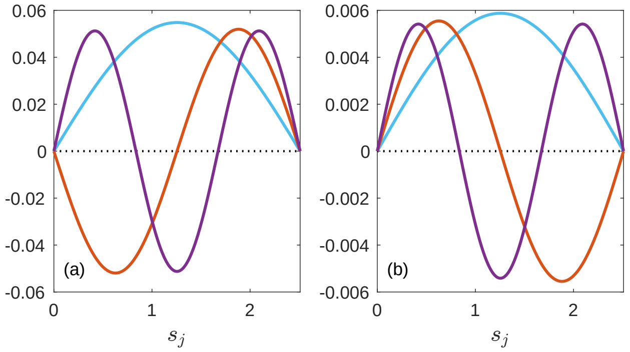


Figure 4. *Modes shapes belonging to each of the dispersion curves in Fig. 3(a).* — $f_1 = 28.7$ *Hz (a) and* 32.7 *Hz (b).* — $f_2 = 57.3$ *Hz (a) and* 65.4 *Hz (b).* — $f_3 = 86.0$ *Hz (a) and* 98.1 *Hz (b).*

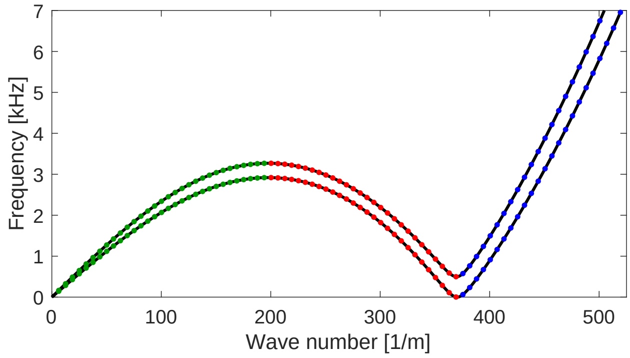


Figure 5. *Modes plotted as dots in the frequency-wave number domain. The solid black lines indicate the dispersion relation curves.*

from the Olson X-82 amplifier [5], now using only one fifth of the length and evaluating over a low frequency range to better visualise the mode frequencies lying on the curve. The three colours split the two mode groups based on Fig. 3(a) using $\beta^-_{\mathrm{T}1}$ and $\beta^-_{\mathrm{T}2}$, and $\beta^+_{\mathrm{T}1}$ and $\beta^+_{\mathrm{T}2}$ for the lower and higher curve, respectively. Note that this grouping could be employed for any set of boundary conditions utilising the more general frequency/wave-like number domain.

Exploring the four excitation/pick-up choices laid out in Section 2.2, the consolidated mode amplitudes in (55) are shown against wave number in Fig. 6, where the colour coding matches that in Fig. 5. Additionally, the same parameters as those in Fig. 5 with the much shorter spring are employed such that the individual modes and their polarity are better visualised. The sub-grouping proposed in Fig. 5 can provide new perspectives when now shown with mode amplitudes. E.g., the choice of excitation/pick-up is shown to cause significant differences in magnitude between the first two sub-groups (green and red). In future work on a fully modelled system, this could provide insights into the significance of various mode groups.

4.4 Simulated Spring Response Decomposition

Following [11], defining a sampling frequency $f_s = 48$ kHz (and corresponding time step $\Delta_t = 1/f_s$) facilitates derivation of a modal update equation from (52):

$$\boldsymbol{\xi}^{n+1} = \boldsymbol{\Gamma}\boldsymbol{\xi}^n - \boldsymbol{\xi}^{n-1} + \mathbf{c}F^n_{\mathrm{E}}, \tag{56}$$

where n is the sample index and $\Gamma_{i,i} = 2\cos(\omega_i \Delta_t)$ [11]. Since the mode amplitudes have been consolidated, the output for each sample is simply the sum of all the modal displacements $\boldsymbol{\xi}^n$.

The modal algorithm is computed for the same spring parameters as in Fig. 4 to visualise a familiar spring reverb response. Of the four cases in Fig. 6, mode amplitudes are derived from exciting/listening to v, one of the bending polarisations. Setting $N = 800$ computes modes that lie on the dispersion curves until around 13 kHz where aliasing in the middle of the eigenvectors begins. While a higher N would yield a close match over the entire hearing range, $N = 800$ is sufficient when considering the strong damping of higher frequencies in real systems (e.g., as seen in Fig. 1).

The simulated response is shown in Fig. 7 and captures some features of impulse responses well, including the dispersive behaviour of echo patterns. The modelled response does not incorporate losses since comparing mode amplitudes to measurements is out of scope without beads incorporated. Fig. 7 instead aims to highlight the grouping which decomposes the response into separate patterns of echoes. Again, the colour coding matches that in Figs. 5 and 6. Sound examples are available here [1] where, for better audibility, an ad-hoc loss mechanism is imposed on (56) similar to [2], but now with a quadratic dependency on wave number.

It is clear from Fig. 7 that the green, red, and blue sections of the dispersion curve each correspond to distinct sets of echoes. This correlates to discussions in [6], where group velocity $v_{\mathrm{g}} = d\omega/d\beta$ is also considered and observations note that flat regions of v_{g} plotted against wave number correspond to coherent wave propagation. In the present study, the versatility of evaluating the grouping by plotting modes as dots in the frequency-wave number domain lies in its immediate extension to the case of a spring with coupled magnetic beads.

Since a simple excitation/pick-up is employed and damping isn't properly considered, the mode amplitudes are of course not matching well with a typical measured response. However, this decomposition of simulated responses will be valuable for analysis in future research aiming to model the full reverb tank. Exploring decomposed responses for the three other excitation/pick-up options in Fig. 6 indicates that further sub-grouping could be explored. In some instances, the isolated echo patterns exhibited faint secondary chirps falling between the strong patterns shown here.

5. CONCLUSIONS

To conclude, isolating several mode groups here has shed new light on understanding modelled spring responses. Distinct sets of echoes have been disentangled in a simulated

[1] https://github.com/jacobmcquillan/SMAC2023-spring-reverb-audio

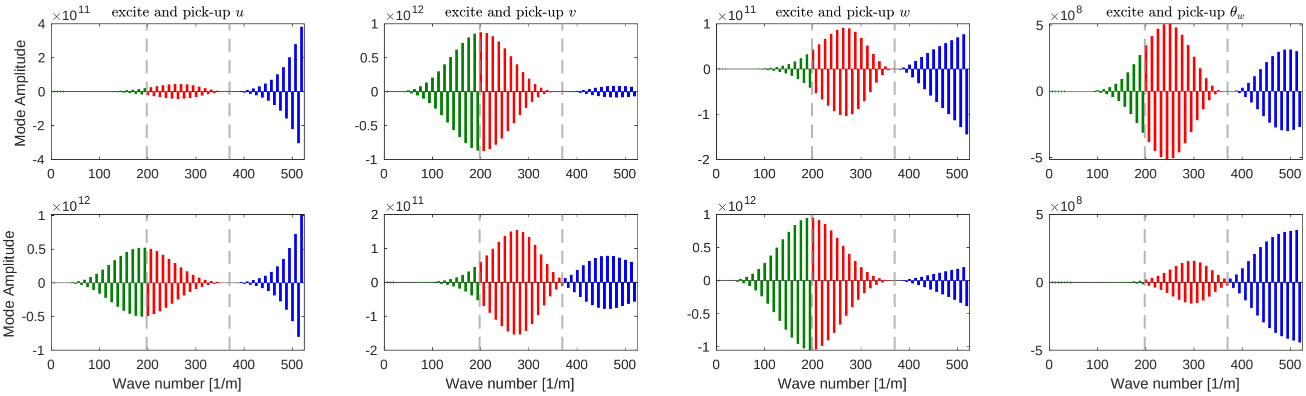


Figure 6. *Consolidated mode amplitudes for four different choices of the driving/pick-up. The top row shows the group lying on the lower dispersion curve with grey lines denoting* $\beta^-_{\mathrm{T}1}$ *and* $\beta^-_{\mathrm{T}2}$*, and the bottom shows the group lying on the upper curve with grey lines denoting* $\beta^+_{\mathrm{T}1}$ *and* $\beta^+_{\mathrm{T}2}$.

response, although this preliminary grouping may need further sub-grouping as some excitation/pick-up choices yield isolated sets of echoes still with elements of faint secondary chirps. Note that this grouping can be employed for any set of boundary conditions, other lossless conditions were not explored here as they are not trivially derived for the model. The grouping proposed here is envisaged to be a useful tool in future analysis when magnetic beads have been incorporated.

Steps have also been taken in the direction of modelling a full reverb tank with stepped beams coupled to a spring. Firstly, an alternative form of the thin spring model allows reduction to a beam in the case of vanishing curvature that retains longitudinal components. Such a system facilitates full exploration of the complex coupling between bending, longitudinal, and torsional vibrations of both the bead and spring.

Secondly, the Chebyshev pseudospectral approach employed here performs reasonably well in audio contexts, looking at a much higher N than in other contexts where only the first few modes are of interest. As such, this method is a viable alternative to higher order finite difference schemes and the increased node density at boundaries may prove advantageous when exploring systems with several connected components. Furthermore, the global nature of pseudospectral approximations avoids entirely ghost nodes that arise in higher order finite difference schemes whose elimination would likely be problematic for these systems. Since N must be quite high here to get a reasonable match in audio contexts, the dense ill-conditioned matrices of pseudospectral methods can pose issues if N was to be pushed further (as would be necessary, e.g., for bigger reverb tanks). To avoid such issues, preconditioning methods may need to be considered [19], or multiprecision computing [2] could be employed since generating modal parameters is independent to running the modal engine.

Acknowledgments

The main author is funded through the Northern Ireland Department for the Economy (DfE) PhD studentship.

[2] https://www.advanpix.com

6. REFERENCES

[1] L. Hammond, "Electrical musical instrument," US Patent No. 2230836, 1941.

[2] J. McQuillan and M. van Walstijn, "Modal spring reverb based on discretisation of the thin helical spring model," in *Proc. of the 24th Int. Conf. on Digital Audio Effects (DAFx20in21)*, 2021, pp. 191–198.

[3] J. McQuillan, M. van Walstijn, J. D. Parker, and M. Ortiz, "Experimental investigation of the driving mechanism in spring reverberation tanks," in *Proc. of Meetings on Acoustics*, vol. 49, 2022.

[4] J. S. Abel, D. P. Berners, S. Costello, and J. O. Smith, "Spring reverb emulation using dispersive allpass filters in a waveguide structure," in *Proc. of the 121st A.E.S. Convention*, 2006.

[5] J. Parker and S. Bilbao, "Spring reverberation: A physical perspective," in *Proc. of the 12th Int. Conf. on Digital Audio Effects (DAFx-09)*, 2009.

[6] S. Bilbao and J. Parker, "Perceptual and numerical aspects of spring reverberation modeling," in *Proc. of 20th Int. Symp. on Music Acoustics*, 2010.

[7] T. Tarnopolskaya and F. de Hoog, "Asymptotic analysis of the free in-plane vibrations of beams with arbitrarily varying curvature and cross-section," *J. of Sound and Vibration*, vol. 196, no. 5, pp. 659–680, 1996.

[8] W. H. Wittrick, "On elastic wave propagation in helical springs," *Int. J. of Mech. Sci.*, vol. 8, no. 1, pp. 25–47, 1966.

[9] S. Bilbao, "Numerical simulation of spring reverberation," in *Proc. of the 16th Int. Conf. on Digital Audio Effects (DAFx-13)*, 2013.

[10] S. Bilbao and J. Parker, "A virtual model of spring reverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 4, pp. 799–808, 2010.

[11] M. van Walstijn, "Numerical calculation of modal spring reverb parameters," in *Proc. of the 23rd Int. Conf. on Digital Audio Effects (DAFx2020)*, 2020.

[12] L. N. Trefethen, *Spectral Methods in Matlab*. Society for Industrial and Applied Mathematics, 2000.

[13] J. P. Boyd, *Chebyshev and Fourier Spectral Methods*. Dover Publications, 2001.

[14] B. Fornberg, *A Practical Guide to Pseudospectral Methods*. Cambridge University Press, 1998.

[15] J.-P. Berrut and L. N. Trefethen, "Barycentric lagrange interpolation," *SIAM Review*, vol. 46, no. 3, pp. 501–517, 2004.

[16] T. A. Driscoll, N. Hale, and L. N. Trefethen, *Chebfun Guide*. Pafnuty Publications, 2014.

[17] L. N. Trefethen, *Approximation Theory and Approximation Practice*. Society for Industrial and Applied Mathematics, 2013.

[18] F. Zhang, *The Schur complement and its applications*. Springer Science & Business Media, 2006.

[19] S. Olver and A. Townsend, "A fast and well-conditioned spectral method," *SIAM Review*, vol. 55, no. 3, pp. 462–489, 2013.

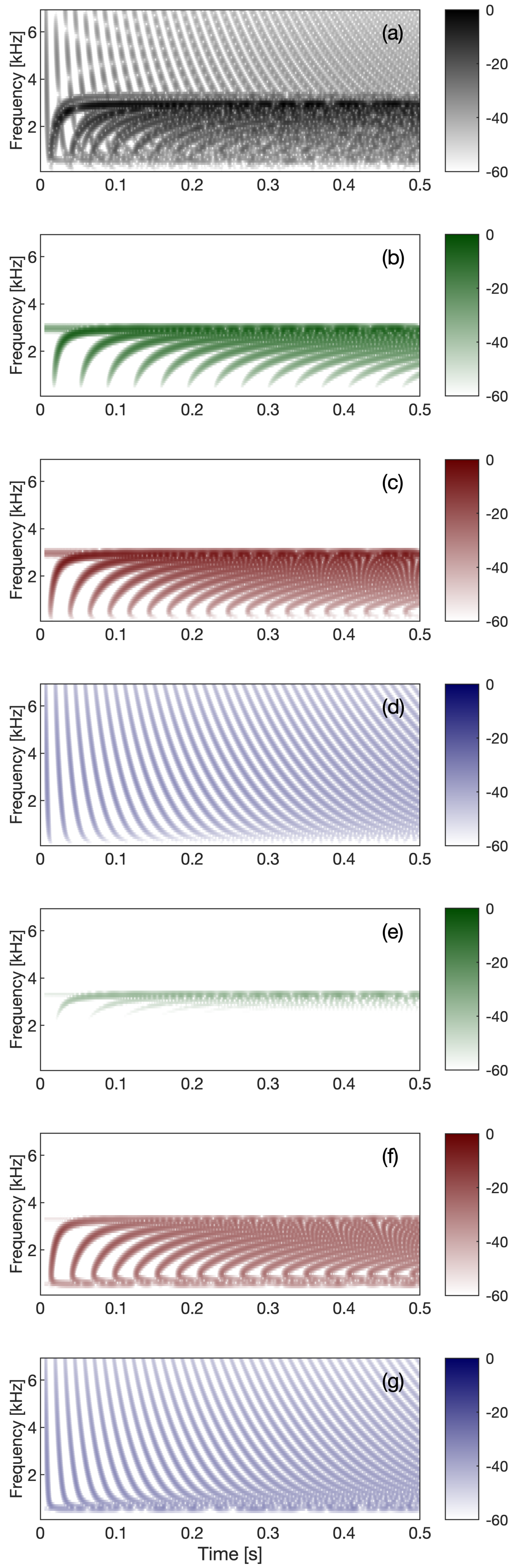


Figure 7. *Simulated spring response (a) split into separate groups belonging to the lower (b–d) and upper (e–g) dispersion curves in Fig. 3(a).*

Estimation of the elastic properties of a violin top plate using finite element models

Mauricio Rodríguez
Undergraduated student,
Escuela de Laudería INBAL,
México
mauricio.rod.munoz@gmail.com

Josie Moran
Undergraduated student,
Escuela de Laudería INBAL,
México
josie.moranm@gmail.com

Inti Wiarco
Undergraduated student,
Escuela de Laudería INBAL,
México
intiwiarco@gmail.com

Patricia Guevara
Undergraduated student,
Escuela de Laudería INBAL,
México
pattyto235@gmail.com

Jorge Vargas
Undergraduated student,
Escuela de Laudería INBAL,
México
jorgevargasrueda8@gmail.com

Jesus A. Torres
Laboratorio de Acústica, Escuela de Laudería INBAL, México
jesusalejandrott@yahoo.com.mx

ABSTRACT

The mechanical properties and geometry of violin top plates determine their vibrational behaviour. In this research, a finite element code simulation of the Stradivarius Titian violin top plate was modified according to the thickness and density of another experimental violin top plate made by the authors. Following, the simulation's mechanical properties were iteratively adjusted until reaching the desired vibratory modes of the experimental top plate. The elastic properties obtained in this way are in agreement with the expected values for spruce, so they were considered a good estimation of the real values for the experimental top plate. The goal of this research is not to replace traditional tuning and making for violin top plates, but to demonstrate how science can work towards arts being an important tool for students and professional violin makers.

1. INTRODUCTION

In violin making, the design of the top plate is very important due the impact in the sound of the instrument. Several studies have been published about the mode shapes of free violin plates. These studies can be labelled in two categories: experimental procedures, where real top plates are tested; and numerical methods, where simulations are performed using computers. Also, there are studies where both procedures are used together.

One of the main advantages of using a numerical model to study a vibratory system is that the model can be quickly modified to predict its response without the need to make any structure. For example, Rogers and Masino (1990) used the finite element method to study a violin top plate. This technique allows simulate structures of complicated geometry through coupled equations systematically assembled.

In this report, the code to create a finite element model of the Stradivari violin Titian (Torres Et. al, 2020) was adapted to simulate the vibrational behaviour of a real top plate. ANSYS student (free version), a computational software based on the finite element method, was used for this purpose. The data of the code were modified until matching the values simulated with the real counterpart. In this way, it was tried to estimate the elastic properties of the wood used to make the top plate.

2. WOOD PROPERTIES

One of the main factors that influence on the quality of a musical instrument, such as the violin, is the quality of the wood used to make it. Traditionally, a top plate is made using spruce wooden blanks (*Picea abies*), with two pieces glued in long grain direction. Low-density wood is preferred by violin makers. They knock the wood to hear the sound produced and therefore, evaluate the quality of the material. However, a deeper knowledge of the wood is required since it will be involved in dynamical behaviours (Limón, 2022).

The mechanical properties are the parameters responsible to define the vibratory response of any structure, as the resonant frequencies and mode shapes. There is a consensus (Hutchins, 1981) about the most relevant properties in the vibration of wood plates: longitudinal and tangential elastic modulus, the main plane shear modulus, damping, and density. Also, measurements such as the sound velocity through the material combine some of these properties, obtaining a more manageable single value.

Elastic and shear modulus can be linked to specific mode shapes in a free violin plate. The second and fifth free mode shapes are useful to analyze this link since the second tends to have longitudinal nodal lines while the fifth nodal lines cross in the tangential direction of the wood. In fact, both modes have been widely studied in the literature on violin making.

3. METHODOLOGY

3.1 Chladni method

The Chladni method is an experimental procedure to visualize mode shapes in vibratory plates. This method consists in spreading sand over the plate under study (a violin top plate in the case of the report) while the structure is driven in a resonant frequency of interest. An amplified loudspeaker and a signal generator are enough for this task. Also, it is strongly recommended obtaining the resonant frequency from a spectrum analyser recording an impulse response of the plate.

In this report, the Chladni method was implanted as follows. Visual Analyzer, a free software, was installed to obtain the impulse response of the top plate to detect resonant peaks with the frequencies of modes 2 and 5. Then, the top plate was horizontally supported in three points: two rubbers below expected nodal lines, and a stinger glued on a loudspeaker below an expected antinodal point. Finally, the desired resonant frequency was sent to the loudspeaker using the signal generator also available in Visual Analyzer. Once all the system was working, the sand was scattered (see Fig. 1).

Figure 1. Chladni patterns of the second (left) and fifth (right) mode shape of the free top plate under examination

3.2 Simulation

Commonly simulating implies implementing a model. The model must contain key features, behaviours and functionalities of the physical system required for the simulation (Banks Et. al, 2001). Models are theoretical developments, without real existence, referred to as objective schemes. To compare experimental procedures with the properties of the models and the systems, a quantitative exploration of the system must be performed supported with mathematical procedures (Figueroa Et. al, 2014).

Often models are created using numerical simulations. They employ a computer to calculate the solution of a set of equations to estimate a desired behaviour. Once a simulation has the required agreement for the case under study, it can be adjusted to analyse its most relevant properties, seeing how the changes alter the results. Numerical simulations are useful to predict how the real system could behave even before its existence.

The finite element method is a numerical technique to solve problems expressed as set of coupled differential equations. The method is particularly useful in solving systems of complicated geometry, submitted to loads, boundary conditions and/or material properties without analytical solution (Logan, 2012). The method works by dividing the system to be solved into elements and expressing the variables in terms of approximated functions for each element. Finite element simulations can be employed to find problems in vibrations and design, as well as to predict structures' responses even before being made and tested. A very common practice for finite element applications is modal analysis.

ANSYS is the world leader in commercial finite element software. In this report, the ANSYS 2022 R1 Student version (free) was used. The modal analysis performed lasted only 1 minute to deliver the results. The model used was created through commands in a text file, previously created in the Laboratory of the authors. It can be freely downloaded for anyone interested (Torres et al. 2020). Commands in a text file to do the finite element modal has the advantage that it can be quickly adapted, with just a bit of practice. Variations in the model allow including or excluding selected parts of a violin soundbox, as well as modifying all the material properties of each one.

The text code includes variables to control the model, and the commands to create all the simulations. Unwanted parts of the simulation must be turned off, and only the top plate, bassbar, and f-holes must be turned on. It is easily controlled by imposing a 0 or 1 value depending on the case. The thickness of the plate can be adjusted using specific variables for this task, according to Fig. 2: variables in the text code correspond to the volumes delimited by lines in the image of the model.

The code is accessed from the File Menu, in the option "File – read input from …". The modified code is selected and OK button must be pressed. Once the window "solution is done!" window appears, the Postprocessor tree must be expanded to locate the "Results viewer". To plot the results, clicking "DOF solution – displacement of vectorsum" located in "show result item". Then, the Plot Result icon is pressed and the mode to be explored can be selected from the slider.

To find the values for the material properties of the real top plate, the density of its wood was obtained from a rectangular bar of the same sample and the thickness was measured at several points of the top plate. These real values are substituted on the default material properties of the text code. Then, a modal analysis was simulated iteratively, adjusting the longitudinal and tangential Young modulus, Ey and Ex respectively; and the shear modulus of the main plane, Gxy. The aim was to reach the closest values for the simulated frequencies of modes 2 and 5, in comparison with the equivalent values obtained from the real top plate.

```
! Top plate thickness viewed from inside
!     bassbar                         soundpost
!       II
!       II                                O
V6=.0024
                          V5=.0024
 V19=.0024
                         V22=.0024
  V20=.0024
                        V23=.0024
   V21=.0023
                       V24=.0023
        V3=.0026
         V13=.0026
          V14=.0023
           V15=.0024
                    V16=.003
                     V17=.0024
                      V18=.0024
                       V4=.0024
   V9=.0024
                       V12=.0024
  V8=.0024
                        V11=.0024
 V7=.0024
                         V10=.0024
V1=.0024
                          V2=.0024
```

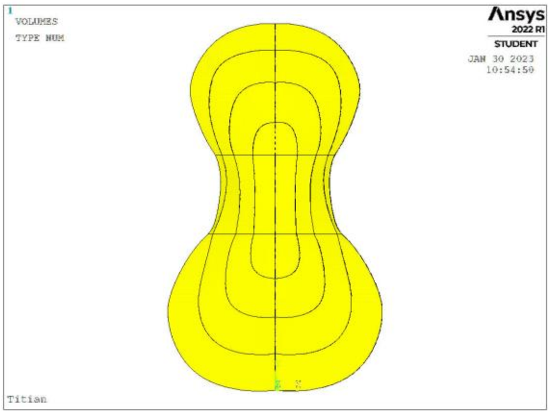


Figure 2. Variables for the thickness in the top plate and the correspondence for each value with the volume in the model. Image courtesy by Ansys Inc.

4. RESULTS AND DISCUSSION

The nodal lines in the mode shapes simulated using the finite element method agree with the experimental results (see Fig. 3). However, matching the values of the simulated resonant frequencies for both modes, at the same time, required considerable changes in the default material properties of the code. The frequencies obtained in both methods are available in Table 1. They were calculated by Ey= 11000 MPa, Ex= 700 MPa, Gxy=4500 MPa and ρ= 416 kg/m^3.

Table 1. Frequencies for free modes of a finite element simulation based on the real top plate measured in this work.

	Experimental	Simulated
Mode 2	162 Hz	178 Hz
Mode 5	330 Hz	321 Hz

It is well known that the vibrational behaviour of a violin, responsible for the sound radiated, is dominated by the mechanical properties of its materials. Therefore, by estimating these values of the wood employed, we can try linking the characteristics of a particular wood sample with the sound quality delivered by the instrument. Of course, estimating the material properties of the wood is not the only ingredient needed to know the specific details of a violin. However, it is hard to find this kind of information when a historical instrument is under examination, so the technique employed in this report could be useful as a tool for this task.

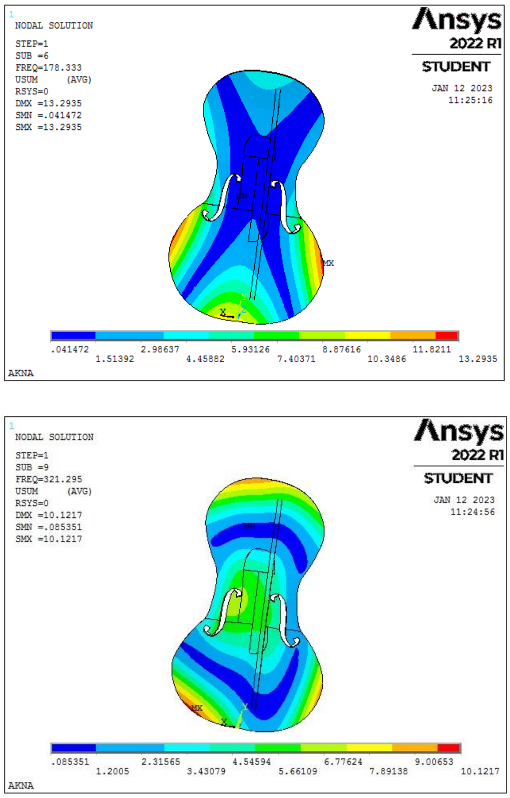


Figure 3. Finite element simulation of the two free mode shapes of the top plate under examination. Images courtesy by Ansys Inc.

5. CONCLUSIONS

The aim of this work about offering a tool to examine material properties in a violin top plates was successfully achieved. Besides, another interesting detail was detected. The procedure here developed can be used as an introduction to finite element simulations. Surprisingly, the last author has found that violin makers have an advantage when they are knowing simulations of the violin, even in comparison with students on physics and engineering. Violin makers *need* a deep understanding of mechanical vibrations and it seems an extra motivation even without a desirable strong background in mathematics.

6. REFERENCES

O. Rogers, T. Masino, "The effect of wood removal on bridges frequencies" Catgut. Acoust. Soc. J, 1990, pp. 1:6-10.

J. Torres, C. Soto, and D. Torres, "Exploring the sign variations of the Titian Stradivari violin using a finite element model". In J. of Acoustical Society of America, 2020, pp. 1499-1505.

D. Limón, "The Effect of the design and wear of the bass bar in a free violin plate" Thesis, Escuela de Laudería, 2022.

C. Hutchins, "The acoustic of violin plates" Scientific American, Vol. 245, No 4, pp. 170, 1981

J. Banks, J. Carson II, B. Nelson, D Nicol, "Discrete event system simulation" Prentice Hall, 2001.

M. Figueroa, S. Ramírez, J. Alcalá, "Introduction to de Experimental Method" PEARSON México, 2014.

D. Logan "A first course in the finite element method" Cengage Learning, 2012.

Characterization of Single-Reed Instrument Sound Generation Based on Ffowcs Williams-Hawkings Analogy

Song Wang song.wang5@mail.mcgill.ca
Esteban Maestre esteban.maestre@mcgill.ca
Gary Scavone gary.scavone@mcgill.ca

Computational Acoustic Modeling Laboratory (CAML)
Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT)
Schulich School of Music, McGill University

ABSTRACT

The Ffowcs Williams-Hawkings (FW-H) acoustic analogy is applied to investigate the sound generation of a single-reed instrument. An FW-H formulation is derived using the one-dimensional Green's function for an infinitely long pipe, which estimates the outgoing acoustic pressure at an observer placed inside the instrument via a surface integral of hydrodynamic variables in the mouthpiece-reed system. The FW-H estimation is based on a two-dimensional computational aeroacoustic model developed with the lattice Boltzmann (LB) method, which computes the integrands in the FW-H formulation. The FW-H acoustic analogy is validated by comparing the estimated pressure at the observer to that simulated by the LB model and a good agreement is found. The outgoing pressure at the observer is further decomposed into contributions from monopole and dipole sources, which correspond to different terms in the FW-H formulation. The monopole sources come from the modulated jet flow entering the mouthpiece and the displacement flow induced by the moving reed, whereas dipole sources are produced by the unsteady force exerted on the fluid by the solid walls. Results show that dipole sources, particularly those associated to the long inclined mouthpiece baffle, dominate the observed pressure at the studied playing frequency of 230 Hz.

1. INTRODUCTION

The sound generation of a single-reed instrument is a multiphysics problem that involves acoustics, fluid dynamics and solid vibrations. When a player plays the instrument, the air flows from the lungs toward the instrument, causing the pressure to build up in the player's mouth. The mouth pressure drives the air to flow through the reed aperture at the tip of the mouthpiece into the instrument. In the meantime, the mouth pressure forces the reed to move toward the mouthpiece, and the moving reed in turn modulates the airflow going through the reed channel. The moving reed also creates airflow, and the so-called reed-induced flow disturbs the air in the acoustic resonator together with the pressure-driven flow. The air disturbance travels back and forth in the resonator, and the reflected energy is fed into the mouthpiece-reed system to support the reed oscillation.

The scientific study of single-reed instrument sound generation can be traced back to the 1860s when Helmholtz [1] theoretically investigated the interaction between the reed and the pipe. More recent researchers, such as Backus [2], Benade [3], Nederveen [4], and Worman [5], contributed to the development of a more general mathematical framework for single-reed instrument sound generation. Such a model assumes a localized interaction between the reed, flow, and pipe, which implies that

- the distributed reed vibration is simplified as a single-degree-of-freedom oscillator,
- the distributed airflow and its interaction with the reed is localized at the tip of the reed, and
- the interaction between the resonator and the generator (mouthpiece-reed system) is localized at the entry of the resonator.

This model serves as the basis of sound synthesizers [6] and nonlinear dynamical system studies [7], and it has been widely applied to explore the instrument's sound properties and oscillation.

In addition to this simplified model, the computational fluid dynamics (CFD) and computational aeroacoustic (CAA) models have been applied to investigate the fluid-structure-acoustic interaction in a distributed physical space. These models have shed light on several aspects of sound generation, such as the quasistationary assumption of the fluid model and its dependence on the mouthpiece geometry [8, 9], the effects of the lip on the sound generation [10, 11], and the influences of the mouthpiece geometry on the sound [9, 12]. However, in terms of sound generation, there has been more discussion of aerodynamics than aeroacoustics, and an efficient way to connect the near-field fluid dynamics in the mouthpiece-reed system to the far-field radiated sound characteristics has yet to be established.

This paper aims at a better understanding of the sound generation and a more in-depth characterization of the sound sources of the single-reed instrument. A two-dimensional (2D) computational aeroacoustic model is built for the single-reed instrument using the lattice Boltzmann (LB) method, and the Ffowcs Williams and Hawkings (FW-H) acoustic analogy is employed to estimate the acoustic pressure in the pipe using LB simulation results in the mouthpiece. The FW-H analogy helps decompose the sound source into distributed monopole and dipole contributions, which are attributed to different sound generation mechanisms and different parts of the mouthpiece-reed system. This work is very much inspired by the research in human phonation [13,14], which made use of the Ffowcs Williams-Hawkings acoustic analogy to characterize the human phonation sound generation mechanisms. A similar analysis routine is used here in studying the same problem but for a single-reed instrument.

2. COMPUTATIONAL AEROACOUSTIC MODELING OF THE SINGLE-REED INSTRUMENT

2.1 Lattice Boltzmann Method

The lattice Boltzmann method (LBM) is an alternative to traditional Navier-Stokes solvers for solving computational fluid dynamics problems. It is based on the lattice Boltzmann equation (LBE), which is obtained by discretizing the mesoscopic-scale Boltzmann equation in physical space, velocity space and time. The LBE is given as

$$f_i(\boldsymbol{x} + \boldsymbol{e}_i \Delta t, t + \Delta t) = f_i(\boldsymbol{x}, t) + \Omega_i(\boldsymbol{x}, t), \tag{1}$$

where f_i is the particle distribution function, e_i is the discrete velocity, Δt is the lattice-unit time step that usually equals 1, and $\Omega(\boldsymbol{x}, t)$ is the collision operator that models the redistribution of the particle population after collisions between particles. The D2Q9 model is used in this paper, which discretizes the velocity space into nine directions for a two-dimensional space. The lattice sound speed is defined as $c_s = 1/\sqrt{3}$, the weight coefficients w_i are set as

$$w_i = \begin{cases} \frac{4}{9}, & i = 0, \\ \frac{1}{9}, & i = 2, 4, 6, 8, \\ \frac{1}{36}, & i = 1, 3, 5, 7, \end{cases} \tag{2}$$

and $\boldsymbol{e}_i$ is given as

$$\boldsymbol{e}_i = \begin{cases} (0, 0), & i = 0, \\ (\cos \frac{(i+2)\pi}{4}, \sin \frac{(i+2)\pi}{4}), & i = 2, 4, 6, 8, \\ \sqrt{2}(\cos \frac{(i+2)\pi}{4}, \sin \frac{(i+2)\pi}{4}), & i = 1, 3, 5, 7. \end{cases} \tag{3}$$

The collision is modeled with recursive regularized BGK (rrBGK) [15], which is known to provide greater numerical accuracy and a more stable simulation for high Mach number and high Reynolds number flow by filtering out non-hydrodynamic components in the solution.

Palabos [16], an open-source LBM-based computational fluid dynamics framework, is used for the numerical simulation.

2.2 Mouthpiece Model

The schematic of the simulation setup is shown in Figure 1. The pressure source is placed at the inlet of the computational domain using the absorbing boundary condition (ABC) proposed by Kam et al. [17]. The immersed boundary method (IBM) [18] is used to model the walls of the mouthpiece and reed, as well as their interaction with the surrounding fluid. The IBM is only applied to the complex geometry of the mouthpiece before the throat (the junction between the chamber and the cylindrical parts of the mouthpiece), whereas the Zou-He boundary condition [19] is applied to the rest of the solid walls, including the walls in the cylindrical bore and mouth cavity. The immersed boundary (IB) nodes and Zou-He boundaries are represented by dotted and solid lines in Figure 1, respectively.

The mouthpiece geometry was derived from a CT (computed tomography) scan of a Meyer® 5M alto saxophone mouthpiece with a tip opening of 1.8 mm. The length of the cylindrical part of the mouthpiece, known as the mouthpiece bore, is 1 cm with a diameter of 1.5 cm. A 2D pipe with the same diameter and a length of 30 cm is attached to the mouthpiece. The characteristic-based time-domain impedance boundary condition (C-TDIBC) [20] is applied to the end of the pipe, where an unflanged cylindrical pipe radiation impedance Z_{rad} is employed to represent the radiation domain. The applied radiation impedance is derived from a s-domain polynomial approximation of the radiation coefficients [21], which is transformed to the z-domain using the bilinear transform.

The sound speed is set to $343\,\text{m/s}$. The kinetic viscosity is set to $1.51e^{-4}\,\text{m}^2/\text{s}$, which is an order of magnitude larger than that of the air to guarantee a stable simulation. The grid size and time step are $\Delta x \approx 9.53e^{-5}$ m and $\Delta t \approx 1.60e^{-7}$ s, correspondingly, which guarantees a 360 point-per-wavelength at $10\,\text{kHz}$.

2.3 Reed Model

The one-dimensional (1D) distributed reed model proposed by Avanzini et al. [22] is used in the present study, and the mouthpiece-reed-lip interaction is illustrated in Figure 2. The reed is modeled in a separate coordinate $(x_{\text{reed}}, y_{\text{reed}})$ in the LB domain, as illustrated in Figure 1, and the subscripts are omitted for simplicity in this section.

The reed is modeled as a clamped bar with its transverse oscillation amplitude y governed by the following Euler-Bernoulli equation:

$$\frac{\partial^2}{\partial x^2} [\, YI(x) \left(1 + \eta \frac{\partial}{\partial t}\right) \frac{\partial^2 y}{\partial x^2} \,] + \rho_r S(x) \left[\frac{\partial^2 y}{\partial t^2} + \gamma_B \frac{\partial y}{\partial t}\right] = F(x, t), \tag{4}$$

where ρ_r is the reed density, Y is the Young's modulus, and η is the magnitude of the internal viscoelastic losses. $I(x) = S(x)\kappa^2(x)$ represents the moment of inertia about the longitudinal axis with $\kappa(x)$ representing the radius of gyration of the cross-section $S(x) = wb(x)$, where w and $b(x)$ represent the width and thickness of the reed.

The force applied on the reed $F(x, t) = F_{\text{lay}}(x, t) + F_{\text{lip}}(x, t) + F_{\text{fluid}}(x, t)$ is composed of the contact force due to the collision between the reed and the lay, the lip force distributed over a contact area of the reed with the lip, and the aerodynamic force from the surrounding fluid.

The contact force F_{lay} comprises the elastic force F_{el} and dissipative force F_{dis}, which are defined correspondingly

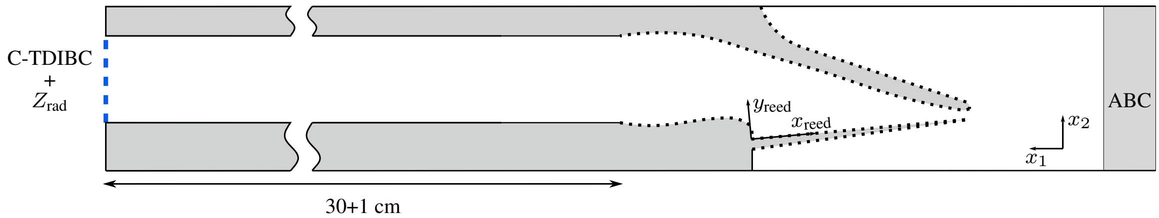


Figure 1. The schematic view of the computational domain.

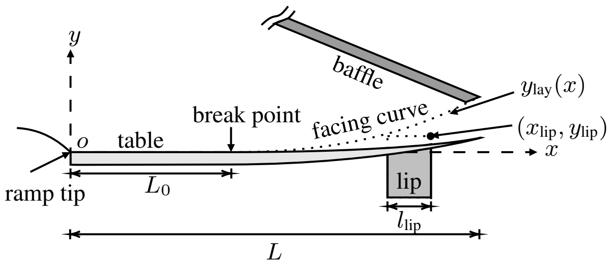


Figure 2. The schematic view of the mouthpiece-reed-lip interaction.

as

$$F_{\text{dis}}(x,t) = \rho_r S(x)\dot{y}(x,t)/\Delta t, \tag{5}$$

and

$$F_{\text{el}}(x,t) = \begin{cases} -K_{\text{lay}}\Delta y_{\text{lay}}(x,t) & \Delta y_{\text{lay}} > 0, \\ 0 & \text{otherwise}, \end{cases} \tag{6}$$

where $\Delta y_{\text{lay}}(x,t) = y(x,t) - y_{\text{lay}}(x)$, and the profile of the lay is fit by a fourth-order polynomial, as

$$y_{\text{lay}}(x) = \begin{cases} \sum_{n=0}^{4} l_n(x-L_0)^n, & x > L_0, \\ 0, & x <= L_0. \end{cases} \tag{7}$$

The lip provides both elastic and damping forces, where $F_{\text{lip}} = -k_{\text{lip}}\Delta y_{\text{lip}}$, with $\Delta y_{\text{lip}} = y_{\text{lip}} - y(x,t) + b(x)$ representing the compression of the lip. The damping effect is included by modifying the damping coefficient γ_B, where

$$\gamma_B = \begin{cases} \gamma_{\text{air}} + \gamma_{\text{lip}}, & x \in (x_{\text{lip}} - L_{\text{lip}}, x_{\text{lip}}), \\ \gamma_{\text{air}} & \text{otherwise}. \end{cases} \tag{8}$$

The time discretization of Equation (4) is consistent with that of the LB simulation. The spatial discretization relies on the discretized immersed boundary nodes of the reed. The distance between nodes is set to $0.45\Delta x$, where Δx is the LB grid size. The IB nodes located at the top of the reed are evenly distributed along the y-axis, and the bottom IB reed nodes are placed based on the thickness function $b(x)$. The reed model has an equal number of IB nodes on the top and bottom, and the F_{fluid} is calculated by subtracting the IB force exerted on the top reed nodes from corresponding ones on the bottom. The IBM is used to update the IB force in the LB simulation at each time step.

3. FFOWCS WILLIAMS-HAWKINGS ACOUSTIC ANALOGY

The acoustic analogy was proposed by Lighthill who reformulated the Navier-Stokes equation into an inhomogeneous wave equation [23]. The nonlinear terms are moved to the righthand side (RHS) of the equation and are considered as the sound source. Ffowcs Williams and Hawkings extended Lighthill's acoustic analogy by introducing moving boundaries [24], resulting in the FW-H equation, written as

$$\left(\frac{1}{c_\infty^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x_i^2}\right)[\rho' c_\infty^2 H] = \frac{\partial^2(HT_{ij})}{\partial x_i \partial x_j} + \frac{\partial F_i}{\partial x_i} + \frac{\partial Q}{\partial t}, \tag{9}$$

where H is the Heaviside function, c_∞ is the speed of the sound in the quiescent flow and

$$\begin{aligned} T_{ij} &= \rho v_i v_j + [p' - \rho' c_\infty^2]\delta_{ij} - \tau_{ij}, \\ F_i &= -(\rho v_i(v_j - \bar{v}_j) + p_{ij})\frac{\partial H}{\partial x_j}, \\ Q &= (\rho v_j - \rho' \bar{v}_j)\frac{\partial H}{\partial x_j}, \end{aligned}$$

and τ_{ij} is the viscous shear stress, v_j and $\bar{v}_j$ represent the velocities of the flow and solid wall, respectively.

The three source terms on the RHS correspond, respectively, to

- the *quadrupole* sound source $\partial^2(HT_{ij})/\partial x_i \partial x_j$ due to the distributed Lighthill stress tensor T_{ij} in the volume,
- the *dipole* sound source $\partial F_i/\partial x_i$ generated by
  - the compressive stress p_{ij} applied to the fluid by the surface, and
  - the momentum flux $\rho v_i(v_j - \bar{v}_j)$ through the surface, and
- the *monopole* sound source $\partial Q/\partial t$ contributed by the mass flux $\rho v_j - \rho' \bar{v}_j$ across the surface.

The solution to the FW-H equation is obtained by convolving the sound source, i.e. the RHS of Equation (9), with the Green's function $G(\boldsymbol{x}, t|\boldsymbol{y}, \tau)$, and the result is stated in the following form after a series of simplifications [25]:

$$\begin{aligned} \rho'(\boldsymbol{x},t)c_\infty^2 = &\int_{-\infty}^{t}\int_V \frac{\partial^2 G}{\partial y_i \partial y_j} T_{ij} \mathrm{d}V(\boldsymbol{y})\mathrm{d}\tau \\ &+ \int_{-\infty}^{t}\int_S \frac{\partial G}{\partial y_i}\left[\rho v_i(v_j - \bar{v}_j) + p_{ij}\right] n_j \mathrm{d}S(\boldsymbol{y})\mathrm{d}\tau \\ &- \int_{-\infty}^{t}\int_S \frac{\partial G}{\partial \tau}\left[(\rho v_j - \rho' \bar{v}_j)n_j\right] \mathrm{d}S(\boldsymbol{y})\mathrm{d}\tau. \end{aligned} \tag{10}$$

The integral surface S and integral volume V for the FW-H formulation in a single-reed instrument are illustrated in Figure 3. The integral surface is composed of the inlet S_{in},

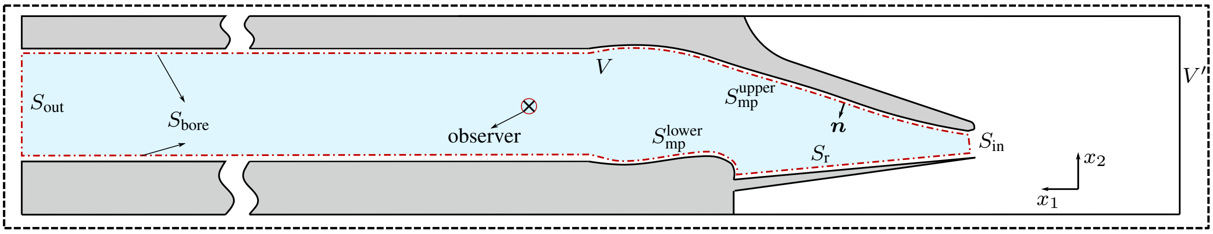


Figure 3. The integral surface of FW-H acoustic analogy with the normals of the surface $\boldsymbol{n}$ pointing into the fluid.

outlet S_{out}, and solid walls $S_w = S_{\text{mp}}^{\text{upper}} + S_{\text{mp}}^{\text{lower}} + S_r + S_{\text{bore}}$ that include the upper $S_{\text{mp}}^{\text{upper}}$ and lower $S_{\text{mp}}^{\text{lower}}$ parts of the mouthpiece, the bore S_{bore}, as well as the reed top surface S_r. The normal $\boldsymbol{n}$ of the integral surface is pointing into the mouthpiece. The integral volume includes only the area inside the instrument and excludes the mouth cavity and radiation domain. In addition, the area in the reed channel (the space between the mouthpiece tip rail and the reed) is not included in the integral volume, and the inlet S_{in} is placed at the end of the mouthpiece reed channel. This is made to eliminate the potential for an unclear definition of the integral surface in the reed channel during the beating of the reed with the mouthpiece.

In the present study, the observer is placed 8 mm away from the mouthpiece throat in the cylindrical bore. It is worth mentioning that in contrast to external flow applications, the observer in single-reed instruments cannot be placed in an acoustic far field due to the presence of standing waves inside the instrument. Therefore, the observer has to be placed in the acoustic near field, which is one of the primary differences between the present FW-H application in single-reed instruments and previous research in human phonation, where the vocal tract was replaced with an infinite pipe without acoustic feedback.

The one-dimensional Green's function for an infinite pipe is used to solve the FW-H equation

$$G(x_1, t|y_1, \tau) = \frac{c_\infty}{2S} H\left(t - \tau - |x_1 - y_1|/c\right), \qquad (11)$$

where S is the cross-section area at the observer.

Based on the integral domain specified in Figure 3, the FW-H formulation is shown as follows:

$$\begin{aligned} \rho'(x_1,t)c_\infty^2 = & \frac{1}{2c_\infty S}\frac{\partial}{\partial t}\int_V [T_{11}]_{t^*}\mathrm{d}V(\boldsymbol{y}) \\ & + \frac{1}{2S}\int_{S_w} [p_{1j}n_j]_{t^*}\mathrm{d}S(\boldsymbol{y}) \\ & + \frac{c_\infty}{2S}\int_{S_r} [\rho_\infty \bar{v}_j n_j]_{t^*}\mathrm{d}S(\boldsymbol{y}) \\ & + \frac{1}{2S}\int_{S_{\text{in}}} [(\rho v_1 v_1 + p_{11}) + \rho c_\infty v_1]_{t^*}\mathrm{d}S(\boldsymbol{y}) \\ & + \frac{1}{2S}\int_{S_{\text{out}}} [(\rho v_1 v_1 + p_{11}) - \rho c_\infty v_1]_{t^*}\mathrm{d}S(\boldsymbol{y}), \end{aligned} \qquad (12)$$

where $t^* = t - |x - y|/c$ is the retarded time.

In this formulation, the first term corresponds to the quadrupole sound source, which arises due to the Lighthill stress tensor in the control volume. The second term represents the dipole sound source contributed by the force exerted on the fluid by the solid walls, while the third term accounts for the monopole sound source produced by the induced displacement flow of the moving reed. Theoretically, the fourth and fifth terms are combinations of the dipole contribution by $\rho v_1 v_1 + p_{11}$ and the monopole contribution by ρv_1. However, these terms can also be treated as equivalent monopoles produced by the mass flux across the inlet and outlet surfaces $\rho v_1 M_1 + \rho v_1^{\pm}$, where $M_1 = v_1/c_\infty$ is the Mach number in x_1-direction, and $v_1^{\pm} = p_{11}/\rho c \pm v_1$ represents the incoming acoustic velocities. It should be noted that both S_{out} and S_{in} are defined perpendicular to the x_1-axis, so that their normal vectors n_j are correspondingly replaced with $(-1, 0)$ and $(1, 0)$ during the derivation.

The above formulation can be further simplified by omitting the quadrupole term, since its magnitude is typically two orders lower than that of the dipole [13]. Additionally, the dipole contribution from the bore S_{bore} is zero, owing to the wall's parallel orientation with respect to the x_1-axis. Consequently, the second to fourth sound source terms are all located upstream of the observer so that they only contribute to the left-going pressure $p^+(x_1, t)$ at the observer's position. On the other hand, the fifth term represents the only sound source located downstream of the observer and contributes exclusively to the right-going pressure $p^-(x_1, t)$.

Given that the present study focuses on sound generation within the mouthpiece, the final FW-H formulation, which evaluates the left-going pressure at the observer using only upstream surface integral, is as follows:

$$\begin{aligned} p^+(x_1,t) = & \frac{1}{2S}\int_{S_{w'}} [p_{1j}n_j]_{t^*}\mathrm{d}S(\boldsymbol{y}) \\ & + \frac{c_\infty}{2S}\int_{S_r} [\rho_\infty \bar{v}_j n_j]_{t^*}\mathrm{d}S(\boldsymbol{y}) \\ & + \frac{1}{2S}\int_{S_{\text{in}}} [(\rho v_1 v_1 + p_{11}) + \rho c_\infty v_1]_{t^*}\mathrm{d}S(\boldsymbol{y}), \end{aligned} \qquad (13)$$

where the total wall surface area is defined as $S_{w'} = S_{\text{mp}}^{\text{upper}} + S_{\text{mp}}^{\text{lower}} + S_r$. The three different terms are referred to as the dipole, reed monopole, and inlet monopole, respectively, in later discussions.

4. RESULTS AND DISCUSSION

The reed and lip parameters are set so that the instrument can play near its first resonant frequency. The mouth pressure increases from 0 to 6000 Pa in 5 ms and remains steady for 45 ms until the end of the simulation.

The LB simulated outgoing pressure at the observer is calculated as $(\tilde{p}+\rho c\tilde{v})/2$, where $\tilde{p}$ and $\tilde{v}$ represent spatially averaged pressure and velocity over the cross-section area at the observer, respectively. Equation (13) is used to compute the FW-H estimation, with the integrands calculated using the LB simulated values on the integral surface.

Figure 4 displays the time-domain comparison between the LB simulated pressure and the FW-H estimation, which shows a good agreement in both the transient and steady-state signals. The spectra of the steady-state signal are calculated using the "period synchronized sampling" technique [26], and are compared in Figure 5. It also shows a good overall agreement with the largest deviation less than 3 dB.

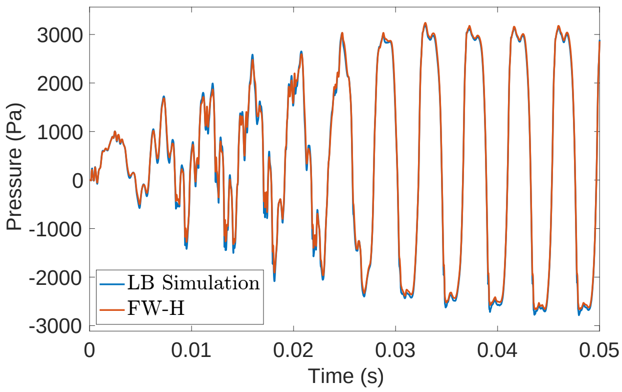


Figure 4. The time-domain comparison between the LB simulated outgoing pressure at the observer and the estimated one using the FW-H formulation.

One of the main benefits of using the FW-H acoustic analogy is its ability to decompose the sound into contributions from various sound generation mechanisms. The equivalent sound level L_{eq} is used to measure the strength of different sound sources:

$$L_{\mathrm{eq}} = 10 \log \frac{1}{T} \int \frac{p(t)^2}{p_0^2} \mathrm{d}t, \tag{14}$$

where p is the sound pressure signal, $p_0 = 20\,\mu\mathrm{Pa}$ is the reference pressure, and T is the period of the signal.

The sound sources are decomposed into the dipole, reed monopole, and inlet monopole, corresponding to the first to third terms in Equation (13), respectively. Their time-domain signals are compared in Figure 6, and the L_{eq} of steady-state signals are 181.2 dB, 153.9 dB, and 164.8 dB, correspondingly. It is clear that the dipole sources, which are generated by the unsteady-force exerted by the solid walls to the fluid, contribute the most to the outgoing pressure at the observer. This seems paradoxical because one may assume the modulated jet flow through the reed channel to be the main source in single-reed instrument sound generation. Such a contrary finding mainly comes from the fact that the fluctuating force that contributes to the dipole in FW-H with the 1D Green's function for an infinite pipe is not only composed of the fluctuation in rotational fluid fields such as the vortices, but also the acoustic fluctuation, which is essentially the acoustic response of the mouthpiece solid walls' to the incoming wave from the resonator. In other words, the dipole contributed by the $\dfrac{1}{2S}\displaystyle\int_{S_{w'}} [p_{1j}n_j]_{t^*}\mathrm{d}S(\boldsymbol{y})$ not only accounts for the sound source due to the presence of solid walls in the fluid, but also for the interaction of reflected sound from the resonator with the solid walls [27], which is related to the role of the mouthpiece as an acoustic resonator [28]. The dipole-dominant feature can also be attributed to the choice of the Green's function. If a one-dimensional Green's function for a semi-infinite pipe terminated at the inlet is chosen, a quadrupole source will emerge from the dipole sources and their image dipoles placed on the other side of the closed end. Because quadrupoles are known to be radiationally inefficient, they might contribute less than the monopole sources. In addition, a new monopole sound source will arise from the outlet, which generates an image monopole that represents the reflection of the incoming acoustic wave by the closed end.

Because a closed reed channel cancels the S_{in}, and therefore the inlet's contribution to the sound at the observer, the inlet monopole contributes almost exclusively to the positive pressure signal. The waveform of the reed monopole is similar to that of the inlet monopole because they are both dependent on reed displacement. The reed monopole contribution, on the other hand, is delayed, which is mainly attributed to the 90-degree lag in phase between the reed velocity and reed displacement, which are correspondingly reflected in the monopole terms of the reed and inlet. Furthermore, the amplitude of the reed monopole is about 10 dB lower than that of the inlet monopole. However, the relative strength of the reed monopole to the inlet monopole should be frequency-dependent. Further investigation is necessary to fully assess their relationship.

The dipole sound source is further decomposed into contributions by different walls, and the upper mouthpiece is discovered to be the strongest dipole, as indicated in Figure 7. This is primarily attributed to the long inclined baffle connecting the end of the reed channel to the mouthpiece throat. The dipole contributions from the reed and lower mouthpiece are comparable in amplitude but out of phase, whereas the reed dipole is roughly in phase with the upper mouthpiece dipole. This is because the upper mouthpiece and reed are oriented in the positive direction of x_1, while the ramped wall of the lower mouthpiece is facing in the opposite direction, resulting in a different sign of the axial normal vectors n_1 when computing the dipole contribution, and hence a 180° phase difference. The distribution on solid walls of the compressive stress $(p_{1j}n_j)$ root-mean-square values is presented in Figure 8, which helps better illustrate the natures of different dipole contributions.

5. CONCLUSIONS

A 2D computational aeroacoustic model was built using the lattice Boltzmann method to investigate the single-reed instrument sound generation. The Ffowcs Williams and Hawkings acoustic analogy is employed to analyze the sound generation mechanisms. It helps decompose the outgoing pressure at the observer in the pipe into contributions by different monopoles and dipoles distributed in the

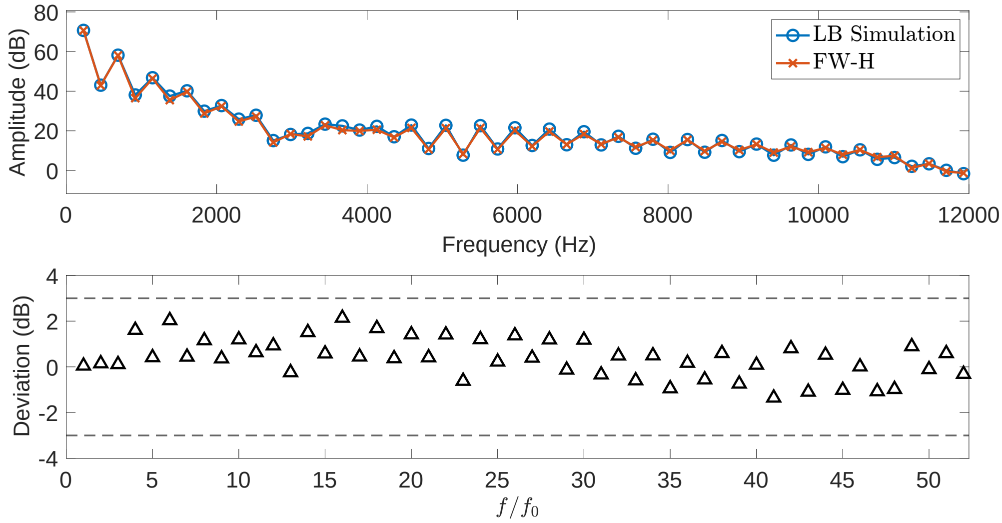


Figure 5. The comparison of the outgoing pressure spectra between the LB simulation and FW-H estimation (top), and the amplitude deviations in dB of harmonics below 12 kHz (bottom).

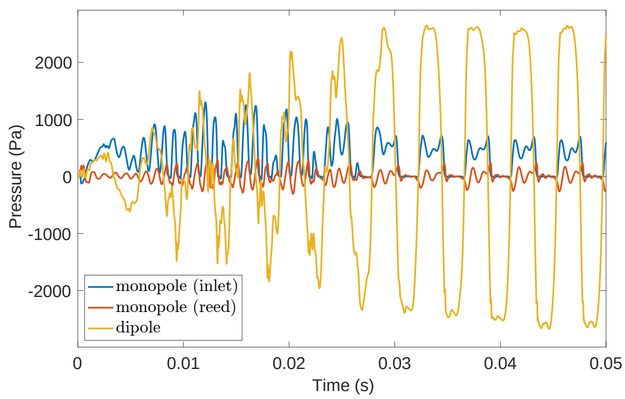


Figure 6. The comparison between different contributions to the outgoing pressure at the observer.

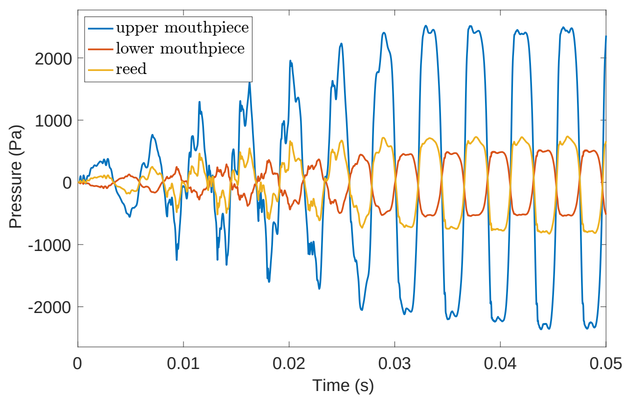


Figure 7. The comparison of dipole sound sources contributed by different solid walls.

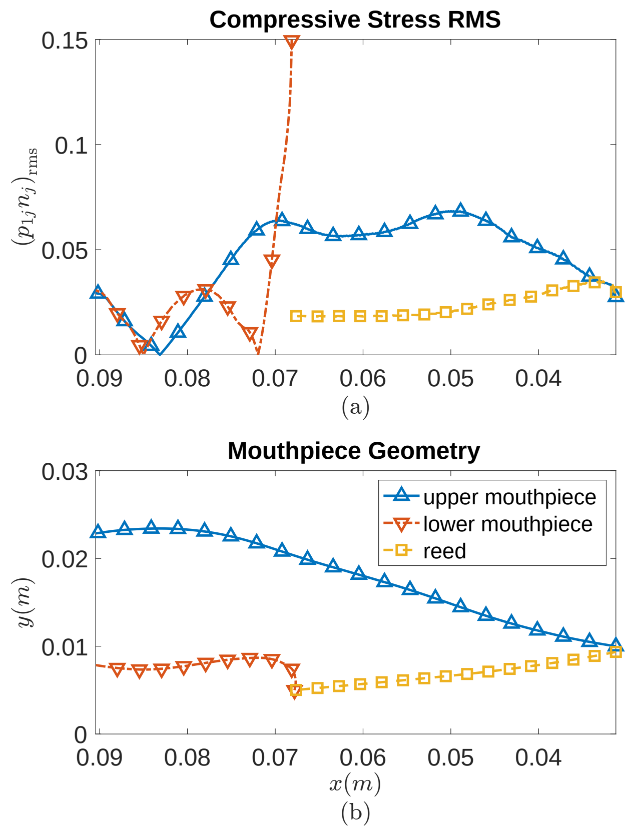


Figure 8. (a) The distribution of $(p_{1j}n_j)_{\mathrm{rms}}$ along the solid walls, and (b) the mouthpiece geometry.

mouthpiece-reed system. The LB simulation results are used to calculate the strength of different sound sources, and the dipole sources, especially the one distributed along the mouthpiece baffle, are the dominant sound sources at the playing frequency considered (around 230 Hz). The FW-H acoustic analogy has been shown to be an effective technique for studying the sound generation of the single-reed instrument. It not only provides additional insights of the sound generation characteristics of the instrument, but also builds a direct correlation between the mouthpiece geometry and the sound, which is useful in mouthpiece design.

The main limitation of the present study is the 2D nature of the computational aeroacoustic model, which lacks the ability to reproduce some of the phenomena found in a real single-reed instrument, such as turbulence and the side slits of the mouthpiece-reed system. However, the present study can be easily extended to a 3D computational aeroacoustic model and use the same FW-H formulation to characterize the sound generation. It is worthwhile to compare the results of a 3D model with those found in this paper, particularly inlet monopole contributions, which are expected to be larger in a 3D model due to the contribution from side slits.

In addition, the dipole-dominant characteristic is partially due to the choice of Green's function. It would be beneficial to derive a FW-H formulation using the low frequency Green's function for a 1D semi-infinitely long pipe terminated at the inlet. This will allow for comparison with the findings presented in this paper, thereby providing further insights into single-reed instrument sound generation mechanisms.

More simulations at different playing frequencies can be studied in the future to characterize the frequency-dependency of different sound sources. Furthermore, it may be interesting to investigate how much the fluid field inside the mouthpiece influences the dipole strength by comparing the present aeroacoustic model with a linear acoustic mouthpiece model, such as the transfer matrix mouthpiece model [28] or transmission line mouthpiece model [29].

Acknowledgments

The authors would like to thank Professor Hirschberg for providing invaluable insights and discussions and the anonymous reviewers for their helpful comments and suggestions. The authors also wish to acknowledge the support from Natural Sciences and Engineering Research Council of Canada, and the Centre for Interdisciplinary Research in Music Media and Technology at McGill University.

6. REFERENCES

[1] H. Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 2nd ed. New York: Peter Smith, 1885.

[2] J. Backus, "Small-vibration theory of the clarinet," *The Journal of the Acoustical Society of America*, vol. 35, no. 3, pp. 305–313, 1963.

[3] A. H. Benade and D. J. Gans, "Sound production in wind instruments," *Annals of the New York Academy of Sciences*, vol. 155, no. 1, pp. 247–263, 1968.

[4] C. J. Nederveen, *Acoustical Aspects of Woodwind Instruments*. Northern Illinois University Press, 1969.

[5] W. E. Worman, "Self-Sustained Nonlinear Oscillations of Medium Amplitude in Clarinet-Like Systems." Ph.D. dissertation, Case Western Reserve University, 1971.

[6] P. Guillemain, "A digital synthesis model of double-reed wind instruments," *EURASIP Journal on Applied Signal Processing*, vol. 2004, pp. 990–1000, 2004.

[7] S. Karkar, "Méthodes numériques pour les systèmes dynamiques non linéaires: Application aux instruments de musique auto-oscillants," Ph.D. dissertation, Aix-Marseille, 2012.

[8] A. R. da Silva, G. P. Scavone, and M. van Walstijn, "Numerical simulations of fluid-structure interactions in single-reed mouthpieces," *The Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. 1798–1809, Sep. 2007.

[9] Y. Shi, "A numerical framework for fluid-acoustic-structure interaction in clarinet-like instruments," Ph.D. dissertation, McGill University, 2016.

[10] N. Giordano and J. W. Thacker, "Navier-Stokes-based model of the clarinet," *The Journal of the Acoustical Society of America*, vol. 148, no. 6, pp. 3827–3835, Dec. 2020.

[11] T. Yoshinaga, H. Yokoyama, T. Shoji, A. Miki, and A. Iida, "Numerical investigation of effects of lip stiffness on reed oscillation in a single-reed instrument," *The Journal of the Acoustical Society of America*, vol. 149, no. 4, pp. A68–A68, Apr. 2021.

[12] ——, "Numerical investigation of effects of mouthpiece geometry on flow and sound generation in a single-reed instrument," in *Proceedings of the 24th International Congress on Acoustics*, Gyeongju, Korea, 2022 Oct.

[13] W. Zhao, C. Zhang, S. H. Frankel, and L. Mongeau, "Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2134–2146, 2002.

[14] C. Zhang, W. Zhao, S. H. Frankel, and L. Mongeau, "Computational aeroacoustics of phonation, Part II: Effects of flow parameters and ventricular folds," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2147–2154, 2002.

[15] O. Malaspinas, "Increasing stability and accuracy of the lattice Boltzmann scheme: Recursivity and regularization," *arXiv preprint arXiv:1505.06900*, 2015.

[16] J. Latt, O. Malaspinas, D. Kontaxakis, A. Parmigiani, D. Lagrava, F. Brogi, M. B. Belgacem, Y. Thorimbert, S. Leclaire, and S. Li, "Palabos: Parallel lattice Boltzmann solver," *Computers & Mathematics with Applications*, vol. 81, pp. 334–350, 2021.

[17] E. W. S. Kam, R. M. C. So, and R. C. K. Leung, "Lattice Boltzmann method simulation of aeroacoustics and nonreflecting boundary conditions," *AIAA journal*, vol. 45, no. 7, p. 1703, 2007.

[18] T. Inamuro, “Lattice Boltzmann methods for moving boundary flows,” *Fluid Dynamics Research*, vol. 44, no. 2, p. 024001, 2012.

[19] Q. Zou and X. He, “On pressure and velocity boundary conditions for the lattice Boltzmann BGK model,” *Physics of fluids*, vol. 9, no. 6, pp. 1591–1598, 1997.

[20] S. Wang, E. Maestre, and G. Scavone, “Lattice Boltzmann Modeling of a Single-Reed Instrument Using a Time-Domain Impedance Boundary Condition,” in *Fourth Vienna Talk on Music Acoustics*, Vienna, Austria, 2022.

[21] F. Silva, P. Guillemain, J. Kergomard, B. Mallaroni, and A. N. Norris, “Approximation formulae for the acoustic radiation impedance of a cylindrical pipe,” *Journal of Sound and Vibration*, vol. 322, no. 1, pp. 255–263, 2009.

[22] F. Avanzini and M. Van Walstijn, “Modelling the mechanical response of the reed-mouthpiece-lip system of a clarinet. Part I. A one-dimensional distributed model,” *Acta Acustica united with Acustica*, vol. 90, no. 3, pp. 537–547, 2004.

[23] M. J. Lighthill, “On sound generated aerodynamically I. General theory,” *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 211, no. 1107, pp. 564–587, Mar. 1952.

[24] J. E. Ffowcs Williams and D. L. Hawkings, “Sound generation by turbulence and surfaces in arbitrary motion,” *Philosophical Transactions of the Royal Society of London A*, vol. 264, no. 1151, pp. 321–342, 1969.

[25] S. Glegg and W. Devenport, *Aeroacoustics of Low Mach Number Flows: Fundamentals, Analysis, and Measurement*. Academic Press, 2017.

[26] T. Grothe, “Experimental Investigations of Bassoon Acoustics,” Ph.D. dissertation, Technische Universität Dresden, 2013.

[27] M. S. Howe, *Theory of Vortex Sound*, ser. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2003.

[28] S. Wang, E. Maestre, and G. Scavone, “Acoustical modeling of the saxophone mouthpiece as a transfer matrix,” *The Journal of the Acoustical Society of America*, vol. 149, no. 3, pp. 1901–1912, Mar. 2021.

[29] S. Wang, G. Scavone, and E. Maestre, “Two methods for acoustic modeling of the saxophone mouthpiece,” *The Journal of the Acoustical Society of America*, vol. 148, no. 4, pp. 2611–2611, Oct. 2020.

Assessment of Performance Characteristics of the Sheng (Chinese Mouth-Organ)

Christopher Clarke **Samuel Goh** **Jer-Ming Chen**
Singapore University of Technology & Design
christopher_clarke@mymail.sutd.edu.sg samuelgoh07@gmail.com jerming_chen@sutd.edu.sg

Da Yang Tan
National University of Singapore
dayangtan@dayangtan.com

Vincent Tan
Singapore Sheng Association (SSA)
bv_tan@hotmail.com

ABSTRACT

The Sheng (Chinese mouth-organ) is a family of free-reed instruments with over 3000 years of documented history and remains an integral member of many traditional East Asian musical ensembles. While the mechanics and acoustic response of individual free-reeds – which may be driven both by inhaling (draw) and exhaling (blow) – have been studied, we describe here the typical performance of the Sheng (as a system of free-reed + pipe resonators) to facilitate further documentation, improvement and propagation of the instrument's design, performance and pedagogy.

To map the typical performance parameter space (SPL vs pitch consistency) of a concert-grade 'alto' Sheng, an expert player sounded the instrument across its tessitura, at three dynamic levels, for both blow vs draw gestures. In terms of SPL, we identify the most efficient register of the instrument, the typical dynamic range available for each note, and the typical consistency expected when delivering a musical dynamic; three types of pressure-signal biases are also identified, regardless of direction of draw/blow. In terms of pitch, we note that (unlike other wind-instruments) intonation tends to flatten with increasing dynamic level. However, when performing the same musical dynamic, intonation is fairly consistent regardless of draw vs blow gestures, with the player delivering the best consistency in the upper register of the instrument.

1. INTRODUCTION

The Sheng – Chinese mouth-organ – is an ancient free-reed instrument extant across East Asia, consisting of a series of pipe resonators, with a free-reed coupled to each pipe, and the reeds share a common pressure reservoir ("wind chamber") which may be driven both by inhaling (draw) and exhaling (blow). While the mechanics and acoustic response of individual free-reeds have been studied [1-3], the typical performance of the Sheng as a system of free-reed and pipe resonators sharing a common pressure reservoir and homogenous design, have not been previously studied under performance contexts.

Here we focus on two performance parameters of a concert-grade "Alto" (Zhongyin) Sheng (Figure 1): sound pressure level (SPL) and pitch deviation (intonation), surveyed across its tessitura (3 octave chromatic range from C3 to B5; 36 pipes), across its full dynamic compass at 3 dynamic levels (*pp*, *mf* and *ff*), and for both breath/pressure directions (blow vs draw gestures). This complements earlier surveys of other Chinese musical instruments [4-6].

Figure 1. Family of six Sheng instruments in contemporary context – the "Alto" (Zhongyin) is second from the right. (Frederic.Lin, CC BY-SA 4.0, Wikimedia Commons)

2. MATERIALS AND METHODS

An expert Sheng player was engaged to play on his 'workhorse' concert-grade "Alto" Sheng. As the expert is a professional member of several traditional musical ensembles, frequent concert soloist, winner of international performance competitions and has over thirty years of performance and teaching experience, and the Sheng used is a concert-grade instrument in frequent use and good working order, we may consider the resulting output to be representative of typical Sheng performance.

In many wind instruments, pitch and SPL (loudness) are inherently interrelated [7-10] so we wish to investigate how these performance parameters interact on the Sheng and understand how difficult (or easy) it is for an expert to deliver 'on-pitch' while assigned a particular musical dynamic level whilst maintaining musicality. Accordingly,

- The player sounded C3 to B5 chromatically (36 notes),
- repeating each note for 4 sets of draw vs blow gestures,
- at three dynamic levels: *pp*, *mf* and *ff* (where *pp* here means "as soft as possible" and *ff* means "as loud as possible" musically, while *mf* was "moderately loud")
- holding each note about three seconds.

In a recording studio treated for noise isolation and absorption, the audio measurements were collected with the player-instrument facing the microphone (Røde NT5) 1 m away. The prevailing SPL is then calibrated against a constant-power excitation source yielding a reference SPL at the position of the player. Pitch and relative sound pressure of each note played (segment of stable pitch identified by ear) were verified and extracted using Adobe Audition CC 2018 to a resolution of 1 cent and 0.1 dB respectively.

3. RESULTS & DISCUSSION

3.1 Types of [AC] Signal Response: Draw vs Blow

For most notes surveyed, when playing at the same dynamic level, a systematic difference between blow and draw gestures can be seen in the pressure signal behavior from the audio recordings, coined as follows:

1. Positive Bias, where the pressure maximum ("max") and minimum ("min") are not symmetrical, but rather drifts in the direction with respect to airflow (i.e. |max| > |min|);
2. Negative Bias, where |max| < |min|
3. Irregular Bias, where |max| ≠ |min| and arise inconsistently within each set of repetitions
4. Zero Bias, where |max| ≈ |min| regardless of direction of draw or blow

Since the microphone's orientation and distance to the Sheng is fixed, the pressure bias may be surmised to arise from differing reed oscillation regimes when blowing and drawing, e.g. asymmetry in reed deflection (or lack thereof). As expected, positive bias is the most common pressure behavior observed for most notes regardless of dynamic level, when accounting for breath direction.

Figure 2 depicts these four bias responses for notes D#3 (positive bias), E3 (zero bias), F3 (irregular bias) and G#3 (negative bias) played *ff*. While across repetitions there may be variance in the overall time-envelope (ADSR), both |max| and |min| remain fairly comparable and consistent when accounting for the direction of airflow.

3.2 Overall *SPL* vs Musical Dynamics

Table 1 and Figure 3 summarizes pitch deviation for all sounded notes measured (C3 to B5 chromatically, repeating each pitch for 4 sets of draw vs blow gestures, played at three dynamic levels: *pp*, *mf* and *ff*).

For most notes, there is a clear separation of sound pressure levels according to the musical dynamic delivered – this is not unexpected, as musical dynamic and perceived loudness are inherently correlated. However, it is gratifying to see the player achieves largely discrete dynamic regions without much overlap – unsurprising, as it is an expert player performing on his concert-grade instrument (the exception is the first octave where particularly C3-D3 have extensive overlap between *ff* and *mf*).

The second octave plays the loudest (54.7±6.5 dB) while the third octave the softest (42.2±8.3 dB; first octave at 51.8±6.4 dB), suggesting this Alto Sheng is designed to be most responsive in the middle of the instrument's tessitura.

Each note on the Sheng has a dynamic range of ~20-25 dB; within that range, each dynamic level (*pp*, *mf*, *ff*) has up to ~7 dB of variability, indicating the extent to which an expert may deliver dynamic control during musical performance – again it is worth bearing in mind that we have optimal conditions: an expert player performing on his concert-grade instrument. It would be interesting to observe the outcome if the player was instead tasked with finer levels of musical dynamics (e.g., *pp*, *p*, *mp*, *mf*, *f*, *ff*).

In terms of blow vs draw gestures, no clear relationships are observed – neither blow nor draw offer systematic influence on output sound pressure and in many instances, they are comparably loud (on average, 49.7±7.2 dB for blow vs 49.5±6.9 dB for draw gestures).

Lastly, we see further examples showing the range of different pressure bias responses with regards to the direct of draw vs blow as already identified in Section 3.1: positive bias (e.g. D#3, F4, G#5 *ff*), negative bias (e.g. D4, D5) or zero bias (e.g. F#3, F#4, B4).

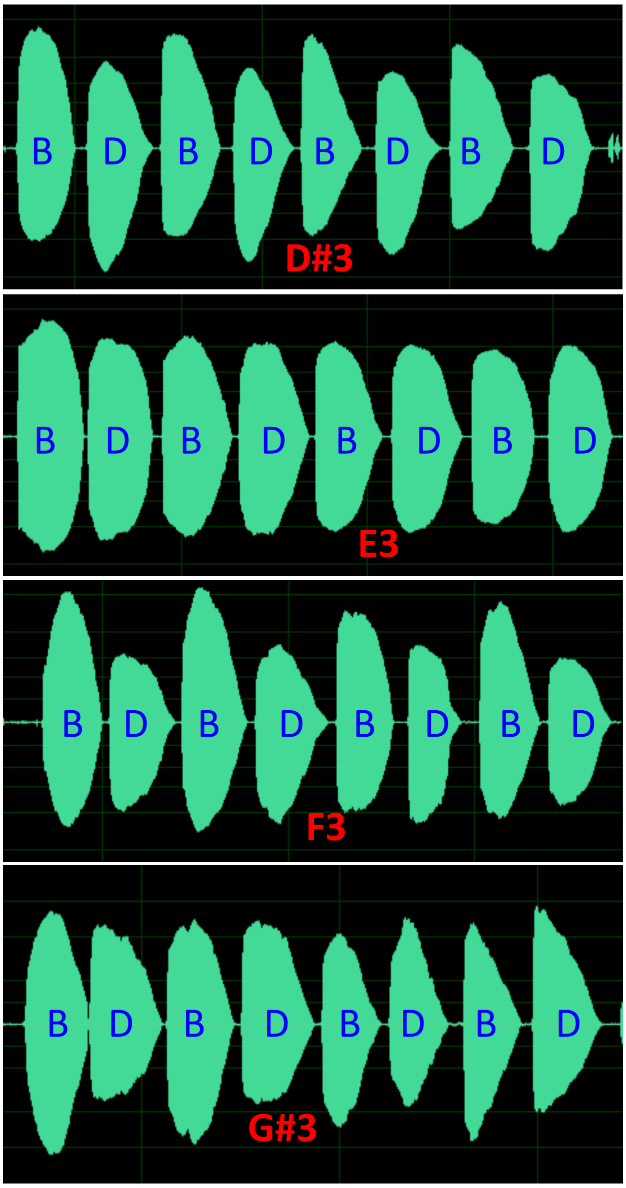


Figure 2. Pressure signals for notes D#3, E3, F3 and G#3 (top to bottom, respective) played *ff* for alternating blowing direction (B – blow; D – draw), showing four types of pressure bias behaviors: positive bias, zero bias, irregular bias and negative bias, respectively (top to bottom).

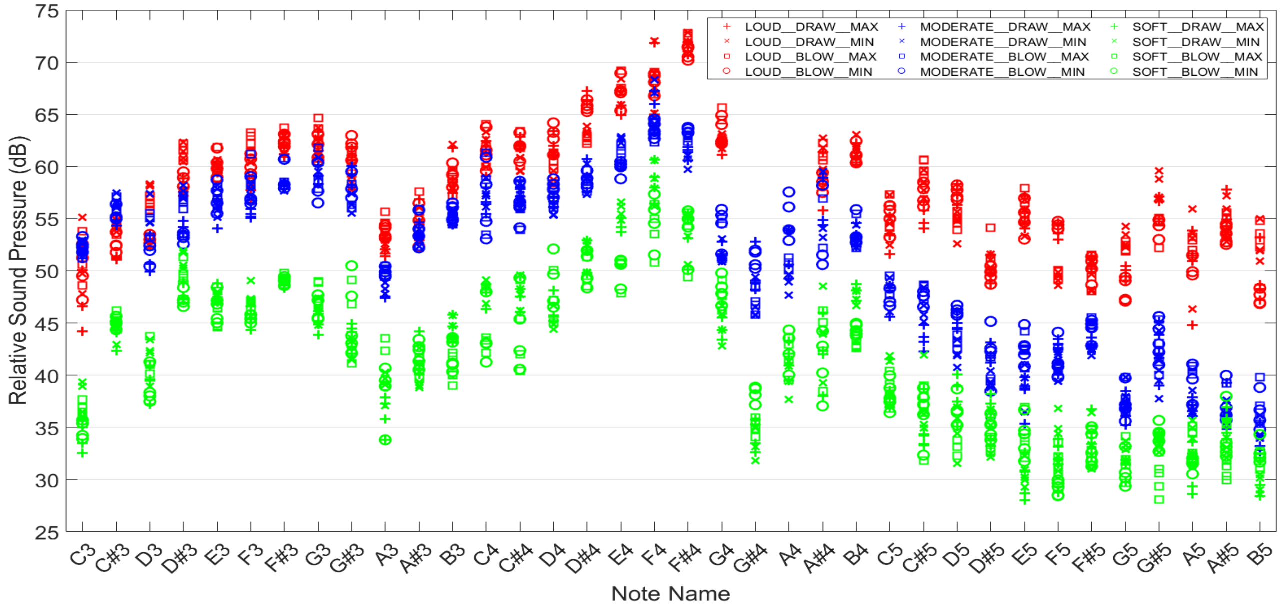


Figure 3. Sound pressure (dB) vs chromatic notes C3 to B5 on the alto Sheng sounded at three dynamic levels (*pp*, *mf*, *ff*) and for four repetitions of draw vs blow gestures. 'Max' and 'Min' refer to the maximum pressure magnitude and minimum pressure magnitude, taking breath direction into account.

SPL (dB)	Draw				Blow				Draw+Blow			
	1st 8ve	2nd 8ve	3rd 8ve	All 8ve	1st 8ve	2nd 8ve	3rd 8ve	All 8ve	1st 8ve	2nd 8ve	3rd 8ve	All 8ve
pp	43.7±1.3	47.5±1.5	33.8±1.7	**41.7±1.5**	43.8±1.4	46.4±2.0	33.6±1.7	**41.3±1.7**	43.8±1.6	47.0±2.4	33.7±1.9	**41.5±1.9**
mf	55.1±1.1	56.4±1.3	40.6±1.4	**50.7±1.2**	55.5±1.2	56.7±1.5	41.9±1.4	**51.3±1.4**	55.3±1.2	56.6±1.7	41.2±1.7	**51.0±1.5**
ff	57.1±1.5	63.9±1.4	53.3±1.8	**57.8±1.6**	57.9±1.3	64.0±1.2	53.2±1.6	**58.1±1.4**	57.5±1.5	64.0±1.4	53.3±1.8	**57.9±1.6**
All Dynamics	**51.6±6.3**	**54.7±6.2**	**42.0±8.3**	**49.5±6.9**	**52.1±6.5**	**52.7±6.7**	**42.4±8.3**	**49.7±7.2**	**51.8±6.4**	**54.7±6.5**	**42.2±8.3**	**49.6±7.1**

Table 1. Summarized SPL accompanying Figure 3, across each octave at 3 dynamic levels and blow/draw gestures.

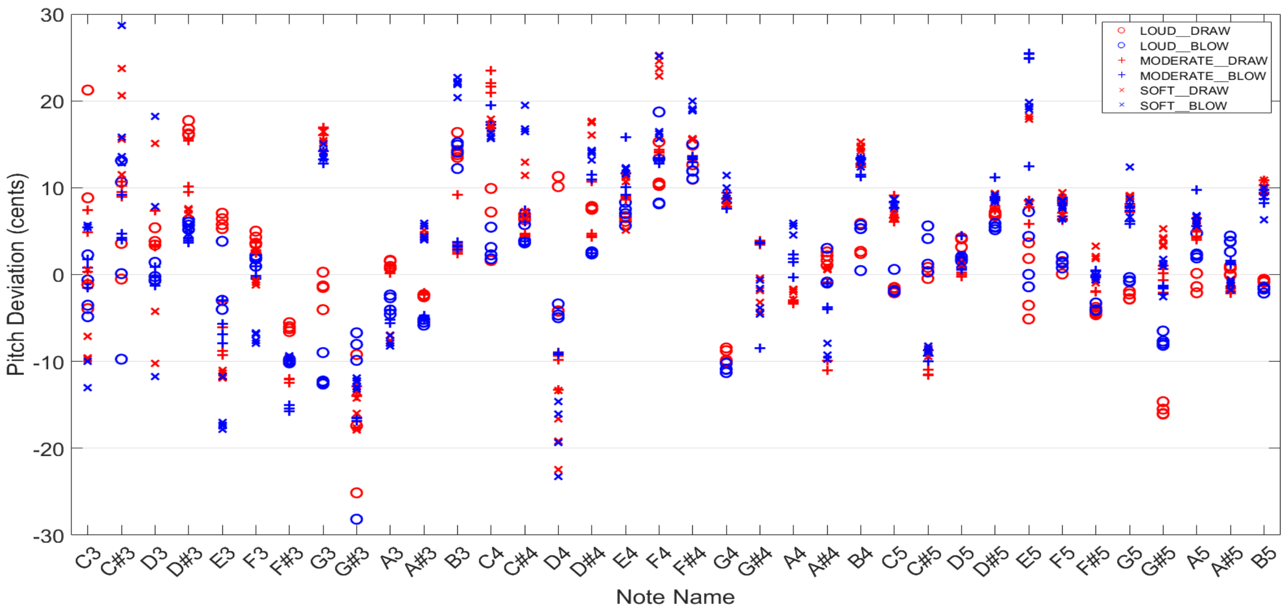


Figure 4. Pitch deviation (cents, reference A4 = 440 Hz) vs chromatic notes C3 to B5 on the alto Sheng sounded at three dynamic levels (*pp*, *mf*, *ff*) and for four repetitions of draw vs blow gestures.

Pitch (cents)	Draw				Blow				Draw+Blow			
	1st 8ve	2nd 8ve	3rd 8ve	All 8ve	1st 8ve	2nd 8ve	3rd 8ve	All 8ve	1st 8ve	2nd 8ve	3rd 8ve	All 8ve
pp	1.0±2.0	7.9±1.1	3.9±3.0	**4.3±2.1**	0.7±2.9	6.9±2.9	4.9±1.1	**4.1±2.3**	0.8±3.1	7.4±3.5	4.4±2.8	**4.2±3.1**
mf	1.6±1.3	6.1±1.0	3.1±0.6	**3.6±1.0**	1.7±0.9	4.6±4.9	5.5±1.5	**2.8±2.4**	-0.0±2.4	5.4±4.5	4.3±2.0	**3.2±3.0**
ff	1.4±4.2	5.0±1.8	-1.1±0.9	**1.6±2.3**	-1.8±2.5	3.1±1.4	0.5±0.9	**0.4±1.6**	-0.2±5.2	4.0±2.3	-0.3±1.6	**1.0±3.1**
All Dynamics	**1.3±6.6**	**6.1±4.7**	**2.1±5.0**	**3.2±5.4**	**-0.9±5.7**	**4.5±7.9**	**3.8±3.8**	**2.5±5.8**	**0.2±6.7**	**5.3±7.0**	**2.9±4.7**	**2.8±6.1**

Table 2. Summarized pitch accompanying Figure 4, across each octave at 3 dynamic levels and blow/draw gestures.

3.3 Overall *Pitch* vs Musical Dynamics

Table 2 and Figure 4 summarizes the pitch deviation (reference A4 = 440 Hz) for all sounded notes (C3 to B5 chromatically, repeating each note for 4 sets of draw vs blow gestures, played at three dynamic levels: *pp*, *mf* and *ff*).

In terms of blow vs draw gestures, there is again no clear difference observed in pitch (on average, pitch differences are <1 cent), except for some notes (e.g. D#3 and A4, where loud draws are ~10 cents sharper than loud blows).

There is good pitch agreement (±6 cents) within most notes, regardless of dynamic level, with the highest octave showing the best overall pitch consistency at ±4.7 cents.

Across dynamics, the typical pitch range expected is Δ12-15 cents, with the second and third octaves having pitch deviating by about Δ10 cents, while broader ranges (Δ20 cents) are seen in the lowest octave. Some notes seem to suffer particularly from poorer pitch stability, e.g. C#3, G3, D4, E5, where pitch instability approaches Δ40 cents.

In terms of 'tonal center' (the local mean pitch), the second octave seems generally sharper, centering around +5 cents, in contrast to the first and third octaves which center around 0 and +3 cents respectively. Given that Section 3.2 identified the second octave as most responsive, one wonders if the added sharpened pitch is intentional to give the instrument perceptual prominence here (to cut through a musical ensemble) at the expense of pitch accuracy?

For most wind instruments, playing at louder musical dynamics is usually associated with a slight sharpening of pitch. However, the reverse is seen here in the Sheng, where the aggregate mean pitch at *pp*, *mf* and *ff* are +4.2, +3.2 and +1.0 cents, respectively (dramatic examples at C#3, G3, F4, B4, C5, C#5, F5 and B5). At the same time, louder dynamics resulting in pitch sharpening can still be seen in some instances (e.g. C3, E3, F3, F#3, D4, D#4, A#4 and C#5). Thus, as both effects are observed and distributed across the instrument's tessitura, we surmise that any dynamic-pitch bias arising here is idiosyncratic for that particular reed-pipe configuration, rather than a function of the instrument maker's deliberate design.

In the second and third octaves, it is also interesting to observe modest local pitch deviations that repeat with octave-level correspondences; for example, between D4-G4 and D5-G5, there is a slight rise in pitch centered around E and F. Although the Sheng is expected to nominally tune for equal-temperament [11], there may be some evidence of a non-equal temperament being favored here, as performers are known to regularly tweak individual reeds for subtle adjustments in response and tuning (but why a more regularized pitch behavior is not seen is also not clear).

4. CONCLUSIONS

We surveyed and report the pitch deviation and relative sound pressure level of a typical concert-grade Sheng with an expert player performing the instrument's standard tessitura at three dynamic levels, outlining and identifying the operating performance parameter space of the instrument.

In terms of sound pressure level, we observe the Sheng to be most efficient at the middle of its tessitura. Each typical Sheng note has a dynamic range of ~20-25 dB and an expert player can deliver each dynamic level to within a range of ~7 dB. No systematic difference between draw and blow gestures are observed, in terms of dynamic levels. Further, we note four types of 'pressure bias' responses are possible, regardless of direction of draw/blow.

Unlike other wind instruments, the Sheng – being a free-reed instrument – does not necessarily show sharpening pitch deviation with increasing dynamic level; instead, the reverse can be observed. Nonetheless, we observe that the middle and upper octave of the Sheng has generally very good pitch consistency independent of dynamic level. Again, no systematic differences in pitch were observed with regards to draw vs blow gestures.

5. REFERENCES

[1] J.P. Cottingham (2002). The Asian free-reed mouth organs. Journal of the Acoustical Society of America, 112(5), 2365-2365.

[2] W. Kuang, P. Leistner, J. Angster, Z. Dubovski, A. Miklos, & J. Yang (2014). Simultaneous detection and analysis of reed vibration and sound of sheng pipes. In Proceedings of the 21st International Congress on Sound and Vibration (pp. 4783-4790).

[3] J.P. Cottingham (2011). Acoustics of free-reed instruments. Physics Today, 64(3), 44-48.

[4] Y.Z. Zhao, S.X. Wu, J.Z. Qiu, H. Huang, H., & L.L. Wu (2010). Sound intensity radiated by gaohu. Science China Technological Sciences, 53(7), 1792–1797. https://doi.org/10.1007/s11431-010-4002-8

[5] N. Li, Y. Zhao, S. Wu, H. Huang, & L. Wu (2012). Sound power level measurement of Chinese bowed stringed instrument - gaoyinbanhu. The Journal of the Acoustical Society of America, 131(4), 3219–3219. https://doi.org/10.1121/1.4708005

[6] J. Qiu, S. Wu, Y. Zhao, H. Huang, & L. Wu (2021). Sound power characteristics of Chinese tradition-al wind instrument Qudi. International Journal of Architecture, Arts and Applications, 7(1), 8. https://doi.org/10.11648/j.ijaaa.20210701.12

[7] J. Wolfe, N.H. Fletcher, & J. Smith (2015). The interactions between wind instruments and their players. Acta Acustica United with Acustica, 101(2), 211-223. https://doi.org/10.3813/AAA.918820

[8] P. De la Cuadra, B. Fabre, N. Montgermont, & C. Chafe (2005). Analysis of Flute Control parameters: A comparison between a novice and an Experienced Flautist. Acta Acustica United with Acustica, 94(5), 740-749. https://doi.org/10.3813/AAA.918091

[9] L. Fuks & J. Sundberg (1996). "Blowing Pressures in Reed Woodwind Instruments", Department for Speech, Music and Hearing. Quarterly Progress and Status Reports (TMH-QPSR), 37(3), 41–56.

[10] A. Almeida, D. George, J. Smith, & J. Wolfe (2013). The clarinet: How blowing pressure, Lip Force, lip position and Reed "hardness" affect pitch, sound

level, and spectrum. The Journal of the Acoustical Society of America, 134(3), 2247–2255. https://doi.org/10.1121/1.4816538

[11] C. Wang, J.Y Chow, & S.S. Wong (2020). The Teng Guide to the Chinese Orchestra. World Scientific Publishing Co. Pte. Ltd and The Teng Ensemble Ltd.

Acoustic Impedance Spectrometry of the Dízi (Chinese Transverse Flute): Fabrication and Fingerings

Hsien Han NG
Singapore University of Technology & Design
hsienhan1987@gmail.com

Da Yang TAN
National University of Singapore
dayangtan@dayangtan.com

Teck Seng NG
Asian Music Technology Ltd
tonexplore@gmail.com

Jer-Ming CHEN
Singapore University of Technology & Design
jerming_chen@sutd.edu.sg

ABSTRACT

The Chinese transverse flute Dízi, as a traditional instrument, is made from found organic material – bamboo. By carefully selecting bamboo blanks of ideal geometry and quality, expert makers fashion them into concert-grade instruments with seemingly minimal intervention (apart from boring embouchure- and finger- holes). We collaborated with a master Dízi-maker from Singapore and document the acoustic impedance spectra at various stages of fabrication, beginning with simple bamboo blanks, until they become completed instruments. These measurements are interpreted in light of geometric, fabrication and material treatment considerations, including observations on the maker's intuitive judgement and subjective assessment of the blanks and their quality. Finally, we present an online compendium of acoustic impedance spectra of one such completed Dízi for all standard (and non-standard) fingerings over the instrument's standard tessitura.

1. BACKGROUND

As a traditional instrument [1], the Dízi's construction as a transverse flute is deceptively simple: a straight section of raw bamboo drilled with an embouchure hole (which the player blows into), a 'membrane-hole' (the non-linear interactions of the driven membrane giving the Dízi its characteristic timbre) slightly downstream, 6 finger holes (keyless per tradition) and a pair of 'foot-holes' (allowing fine tuning adjustments and also to control/homogenize the output timbre across registers). However, because bamboo is a "found" organic material, and for cultural and aesthetic reasons Dízi's tend to feature the bamboo's raw external surface (perhaps with only light polishing or varnishing allowed), the Dízi maker is therefore limited by the natural state of the bamboo blank available: its internal and external dimensions, taper, uniformity, surface conditions, geometric stability, stiffness and density. Consequently, every bamboo blank selected must be approached uniquely and shaped accordingly to optimize its many construction/design parameters/constraints – something a master Dízi maker undertakes by a combination of experience and 'intuition' to produce a top instrument.

To investigate the step-by-step process of Dízi fabrication from raw bamboo stock to completed instrument, we collaborate with a Singaporean master Dízi maker who is regarded by the international Dízi community as creating instruments with superior 'control', 'uniformity', 'responsiveness' and 'playability' [2].

Although acoustics of the transverse flute family have been well studied [3], fabrication is not well explored. Having the opportunity to work with a master Dízi maker, we document (using acoustic impedance spectrometry) and assess the acoustic transformation at each step ("snapshot") of the Dízi fabrication process, provide some observations on the maker's subjective assessment of raw bamboo blank quality, and explore how varnishing may affect the finishing touches to an instrument. We conclude by collating the spectra of all standard (and non-standard) fingerings measured on one such completed Dízi.

2. METHODOLOGY

2.1 Selection of Bamboo Blanks

A set of nine seasoned bamboo blanks (*Pleioblastus amarus*, 苦竹) were selected by the maker to produce "High G Dízi" (i.e. the tonic/'*do*' fingering xxx ooo – see Figure 1 – sounds G5, 784 Hz) and are trimmed to a nominal working length of 366 mm, with characteristic internal and external diameters 14.5 and 21.5 mm respectively at the 'upstream' end and 12.5 and 19.5 mm at the 'downstream' end, exhibiting a gentle taper of −0.08°. (Note: because of the organic nature of bamboo, apart from selecting blanks with these internal and external dimensions ideal for "High G Dízi", no other 'fixed' dimensions are prescribed per se and the maker must judge – for *each* different bamboo blank in hand – a suitable position, size and wall/edge angle of each hole by iteratively playing and adjusting/correcting; a false step cannot be undone!)

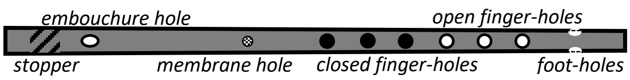


Figure 1. Dízi components, showing fingering xxx ooo for sounding the tonic/'*do*' note (x – closed finger-hole and o – open finger-hole).

Fabrication steps are as follows (*cf.* Figure 1):

- Embouchure hole cut and dressed (i.e. fine adjustments to hole dimensions, wall angles, etc)
- Cork stopper (the obstruction in the bore of the flute, just "upstream" of the embouchure hole) is positioned and the "downstream" end of the tube trimmed to the appropriate length
- Foot-holes cut and dressed

- Finger-holes cut and dressed, beginning at the downstream end (i.e. xxx xxo, xxx xoo, etc)

(Note: because "High G Dízi" plays in the 'piccolo' range, the presence of the membrane hole is optional.)

2.2 Measuring Acoustic Impedance Spectra

To measure the acoustic impedance spectra at each stage of the evolution from blank to finished Dízi, the three microphone two calibration (3M2C) technique [4] is used in a configuration according to previous studies on western classical flutes [5, 6], coupling the flute at the embouchure hole and accounting for the 'missing' radiation impedance at the embouchure hole (associated with the player's lips and face during normal performance) using a short 'correction' spacer (4 mm) introduced between the measurement reference plane and the embouchure hole (*cf.* §3.4, [6]) and sealed using a gasket of adhesive putty ('blu-tac'). The gasket is necessary, because each bamboo blank is neither the same external diameter (at the position of the embouchure hole) nor perfectly cylindrical (exterior surface is natural and unfinished). Thus, to ensure the gasket is well sealed and consistently fitted, each measurement is repeated 5 times, each time by mounting and unmounting the Dízi to the impedance measurement head afresh and the resulting impedance spectra inspected for agreement.

Three Brüel & Kjær 4944A ¼-inch condenser microphones are used in the impedance measurement head, along with Brüel & Kjær Nexus conditioning amplifier, digitized by a MOTU interface, and data collected and analysed in MATLAB. The frequency range chosen was from 100 Hz to 4 kHz, at a frequency resolution of 2.7 Hz.

The measurements were performed in a room partially treated for sound absorption, with ambient temperature 25 ± 0.5°C and relative humidity 55 ± 5% – both lower than playing conditions, so features observed may be expected to deviate very slightly in frequency (~2%; ~30¢) [6].

2.3 Fabrication Steps & Predictions of Quality

For the nine bamboo blanks, it was initially decided that impedance measurements were to be made following each step of their fabrication/evolution process, in order:

0. Blank (trimmed to 366 mm), measured longitudinally "looking in" from the "upstream" end of the tube
1. Embouchure hole cut and dressed, cork stopper fixed, and "downstream" end of the tube trimmed to appropriate length, now measured transversely "looking into" the embouchure hole (as per [5, 6])
2. Foot-holes cut and dressed, thus yielding effectively, the xxx xxx fingering configuration
3. First finger-hole cut and dressed: xxx xxo
4. Second finger-hole cut and dressed: xxx xoo
5. Third finger-hole cut and dressed: xxx ooo
6. Fourth finger-hole cut and dressed: xxo ooo
7. Fifth finger-hole cut and dressed: xoo ooo
8. Sixth and final finger-hole cut and dressed: ooo ooo

However, upon performing completing Steps 1 and 2 on these blanks, our Dízi master maker was already able to *rank* the nine blanks in terms of their potential performance, predicting which ones would yield the best (and worst) Dízi. With this insight, we decided to proceed with Steps 3-8 using only the two best and two poorest blanks, thus yielding the greatest contrast between their performance and corresponding impedance spectra (Section 3.1).

Additionally, one of the 'average' blanks (identified in Step 0) was set aside for a separate study: to examine the effect of varnishing the inner surface. Varnishing the *interior* of the Dízi is an integral process in Dízi making as it imparts the bamboo with protection and stability against moisture and condensation from the player's breath. Varnish was applied on days 3, 4 and 6; impedance measurements were taken just *before* and *after* a fresh coat is applied, and also when allowed to dry.

3. RESULTS & DISCUSSION

3.1 Quality of Bamboo Blanks

Figure 2 compares two good quality and two poor quality blanks, measured in Step 0, by overlaying their *normalized* impedance spectra. Nominally, we see reasonable agreement across these 4 blanks, showing resonance frequencies with almost consistent harmonic distribution and largely similar spectral structure. Despite these similarities, however, it is striking to note that at Steps 0-2, our master Dízi maker had already intuitively and independently assessed and ranked the 9 blanks and identified these exact four blanks as the two best and two poorest performing blanks (across the fabrication and measurement process, he is naïve to the impedance data collected).

Now, because these spectra are normalized (first maximum = 1) and overlaid, a persistent contrast in the resonance minima clearly observed for all resonances shown: the good-quality blanks consistently show both sharper minima (higher Q factor) and lower minima magnitude than the poor-quality blanks. Spectral structure around minima is important because the operating resonance frequencies of flutes – as open-open pipes – are sustained at the impedance minima [5-7] ("maximum flow at minimum pressure"). Thus, holding all else constant, a lower minima magnitude and higher Q factor entails the note will both be easier to sound and sustain, and the resulting pitch should be more stable – factors offering assurances of playability for the player.

Lastly, it may be noted that the spectra of the good-quality blanks show good agreement with each other than the poor-quality blank spectra, which have poorer consistency.

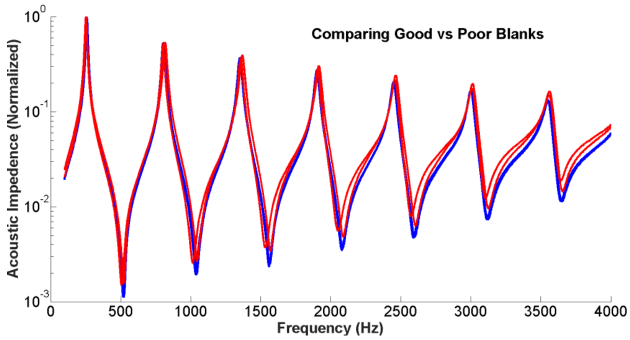


Figure 2. Acoustic impedance spectra contrasting good-quality (blue) vs poor-quality blanks (red).

3.2 Step-by-Step Fabrication "Snapshots"

Figure 3 shows examples of impedance spectra measured at Steps 2, 3, and 4, respectively, for a good-quality blank. These steps correspond to Dízi fingerings xxx xxx, xxx xxo and xxx xoo (Solfège notes '*sol*', '*la*' and '*ti*', respectively), which on "High G Dízi" will nominally sound D5, E5, F#5 using their first resonance seen at 599, 674 and 756 Hz respectively. In each case, the operating resonances (minima) and anti-resonances (maxima) are sharp and clearly defined at lower frequencies but get gradually weaker with increasing frequency as these extrema values converge towards ~3 MPa.s.m^3: the characteristic impedance value corresponding to a bore radius of ~6.6 mm.

Like Figure 2, the resonances in Figure 3 are fairly regularly separated in frequency. For example, the first five resonances of the xxx xxx fingering ('*soh*') are 599, 1202, 1797, 2390 and 2984 Hz, equivalent to sounding pitches D5+34¢, D6+40¢, A6+36¢, D7+30¢ and F#7+14¢ respectively. Consequently, this allows players to overblow notes (overtones) which are nominally in harmonic relationship when using that fingering.

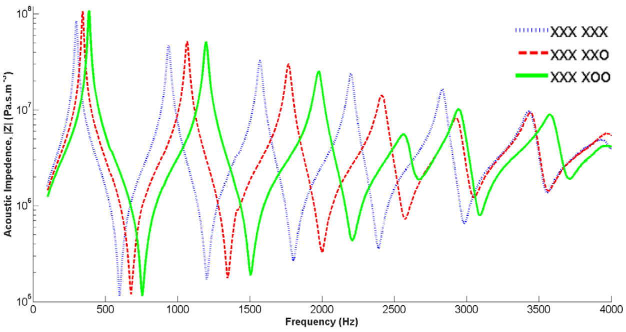


Figure 3. Acoustic impedance spectra measured at Steps 2 (blue), 3 (red), and 4 (green) measured on a good-quality blank, corresponding to fingerings xxx xxx, xxx xxo and xxx xoo (Solfège notes '*sol*', '*la*', '*ti*' notes, i.e. nominally D5, E5 and F#5, respectively, on a "High G Dízi").

3.3 Role of Varnish on Acoustic Response

Daily impedance measurements on the unvarnished blank (from Step 0) on days 1, 2, and 3 demonstrate very good consistency and indicate the acoustic stability of the unvarnished blank, expected of well-seasoned material. Further, Figure 4 shows impedance spectra before application of subsequent coats of varnish on days 3, 4, and 6 – the spectra show remarkable agreement in structure. Although it may be expected that the application of varnish (when 'wet') changes the local porosity of the inner surface of the bamboo pipe (and hence adversely affect sharpness or Q factor of the resonances), we see here that once cured, the application of varnish does not significantly affect or transform the acoustic response of the system per se; inner surface fibers are assumed to remain unaffected (*cf.* [8]).

This means that the choice of raw material (including geometric/dimensional stability, and surface texture) and geometry is crucial right at the outset and varnishing does not seem to offer the Dízi maker a means to 'improve' on poor quality material or instrument geometry. This appears to agree with the views of our Dízi maker: the choice of mediocre bamboo blank limits the outcome of the Dízi quality, and varnishing does not offer redemption; only the best material will do.

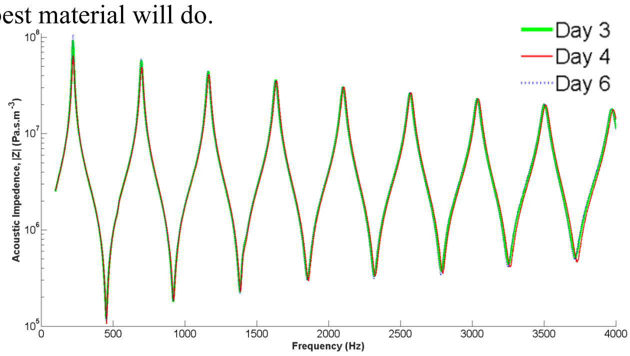


Figure 4. A comparison of impedance spectra just before application of varnish for day 3 (green), 4 (red), and 6 (blue).

3.4 Fingering Database

To conclude this study on the acoustic impedance spectra of the Dízi, 35 standard and alternative Dízi fingerings representing 24 semi-chromatic notes D5 to G7 (the instrument's tessitura) were measured on a finished concert-grade "High G Dízi" by the master Dízi maker. These impedance spectra measurements were then accompanied with a set of audio recording of the same Dízi played using the same fingering, as performed by the master Dízi maker.

This allows insight on how fingerings determine the operating resonances available on the Dízi and show how each fingering can support several resonances and corresponding sounding pitches; conversely, certain target pitches can also be generated using fingering alternatives. These measurements establish a picture of how Dízi players navigate a complex 'map' of fingering possibilities, performance parameters (ease of sounding and pitch control) and musical requirements and elicit insight into how fingering choice, Dízi design and instrument fabrication translate to acoustic response with implications on playability, pitch, and practical performance considerations.

This compendium of Dízi impedance spectra (alongside the sound spectra) is available online [9] (here: https://acoustics.sutd.edu.sg/dizi-impedance/) and complements an earlier database of impedance spectra presented on western classical and modern flutes [10]. Figure 5 offers a screenshot of the online database.

4. CONCLUSION

By making "snapshots" of acoustic impedance measurements accompanying each step of the Dízi fabrication process, we gain some insight on the acoustic implications and design considerations accompanying each fabrication step, as a raw bamboo blank evolves into a concert-grade Dízi. We also corroborated the master Dízi maker's subjective (intuitive) assessment on blank quality by quantifying systematic differences in spectral response between good-quality vs poor-quality blanks, with insights on playability. While varnishing is routinely used to protect the finished instrument from moisture in the player's breath, it neither improves nor compromises the acoustic response of the instrument – a doubled-edged sword. Through careful selection of appropriate bamboo blanks possessing optimal geometric dimensions, material and surface quality, matched with judicious fabrication judgements, good quality Dízi can be created.

Acknowledgments

The acoustic impedance measurements were conducted at the Acoustics Lab, School of Physics, UNSW, Sydney Australia, and the authors are grateful to John Smith and Joe Wolfe for graciously hosting us. The acoustic impedance measurement hardware and software used were developed by Paul Dickens and Noel Hanna.

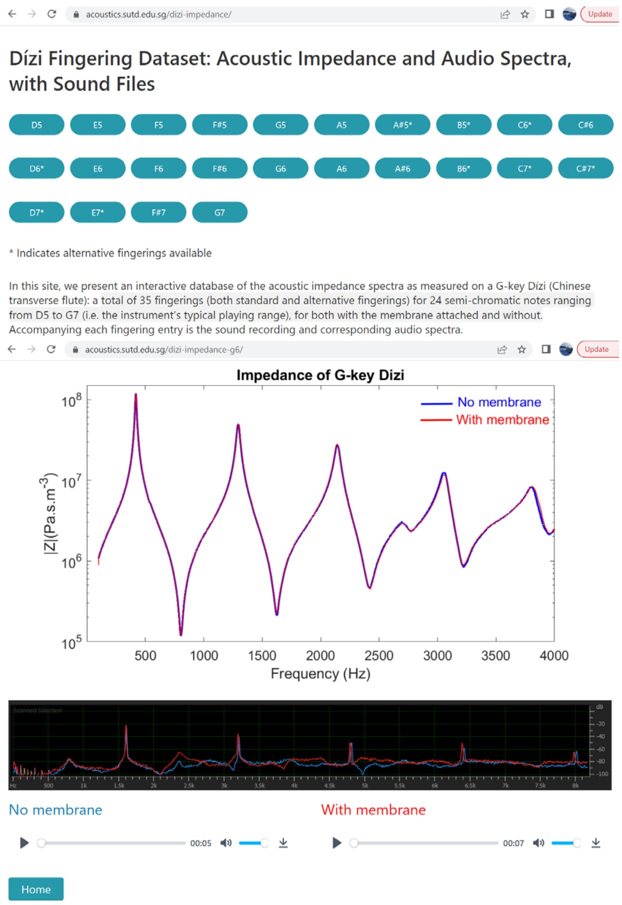


Figure 5. Sample screenshot (Fingering G6) of the online database, showing both the measured acoustic impedance spectra for each fingering, with accompanying audio and audio spectrum.

5. REFERENCES

[1] C. Wang, J. Y. Chow and S. S. Wong, 2020. *The Teng guide to the Chinese orchestra*. Singapore: World Scientific Publishing Co. Pte. Ltd and The Teng Ensemble Ltd.

[2] S. Lam, 2016. In China this Singaporean flute-maker is a 'sifu', *Channel News Asia* feature, https://web.archive.org/web/20160622182634/http://www.channelnewsasia.com/news/singapore/in-china-this-singaporean/2858274.html

[3] N. H. Fletcher and T. D. Rossing, 1998. *The Physics of Musical Instruments*. New York: Springer-Verlag, 503-551.

[4] P. Dickens, J. Smith and J. Wolfe, 2007. Improved precision measurements of acoustic impedance spectra using resonance-free calibration loads and controlled error distribution, *J. Acoust. Soc. America*, **121**, 1471-1481.

[5] J. R. Smith, N. Henrich and J. Wolfe, 1997. The acoustic impedance of the Boehm flute: standard and some non-standard fingerings, *Proc. Inst. Acoustics*, **19**, 315-330.

[6] J. Wolfe, J. Smith, J. Tann and N. H. Fletcher, 2001. Acoustic impedance of classical and modern flutes, *Journal of Sound and Vibration*, **243**, 127-144.

[7] P. Dickens, 2007. *Flute acoustics: measurement, modelling and design*, Doctoral Thesis, UNSW

[8] H. Boutin, S. Le Conte, S. Vaiedelich, B. Fabre & J. L. Le Carrou, 2017. Acoustic dissipation in wooden pipes of different species used in wind instrument making: An experimental study. *J. Acoust. Soc. America*, **141**(4), 2840-2848.

[9] H.H. Ng, Y.X. Tan, T.S. Ng and J.-M. Chen, 2022. "Dizi Fingerings, Impedance and Audio", https://acoustics.sutd.edu.sg/dizi-impedance/

[10] J. Wolfe, J. Smith, J. Tann and N. H. Fletcher, 2001 "Acoustics of classical and modern flutes: a compendium of impedance spectra, sound spectra, sounds and fingerings", https://newt.phys.unsw.edu.au/music/flute/

Detecting efficiency in trumpet sound production: proposed methodology and pedagogical implications

Alberto Acquilino* **Ninad Puranik*** **Ichiro Fujinaga** **Gary Scavone**

Department of Music Research, Schulich School of Music, McGill University
Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT)
555 Sherbrooke St W., Montreal, Canada H3A 1E3

alberto.acquilino@mail.mcgill.ca ninad.puranik@mail.mcgill.ca

ichiro.fujinaga@mcgill.ca gary.scavone@mcgill.ca

ABSTRACT

In this study, we consider the case of the trumpet to study the role of timbre quality from the perspectives of music pedagogy and music information retrieval. Prominent brass pedagogues have reported that the presence of excessive muscle tension and inefficiency in playing by a musician is reflected in the timbre quality of the sound produced, which is easily distinguished by an experienced ear. A technological tool that provides an automatic feedback on the tone quality can be of immense help for new learners to develop good playing habits during independent practice. To develop such a tool, an extensive dataset consisting of more than 19,000 tones played by 110 trumpet players of different expertise has been collected. We manually labelled a subset of 1,711 notes from this dataset with a grade on a scale of 1 to 4 based on the perceived efficiency of sound production. A classifier model with a mean 10-fold cross validation accuracy above 80% was developed based on the extracted audio features to predict the level of efficiency. Finally, we present an interface for the application of this model in pedagogical contexts, in the framework of commercially available music education systems.

1. INTRODUCTION

The use of technological systems aimed at supporting musical instrument learning is a burgeoning area, especially since the spread of portable devices. Educational technologies are designed to assist music students in various aspects of their study, such as checking accuracy of pitch and rhythm [1], fingering [2], musician's form [3], performance [4], and dynamics control [5, 6].

* Equal contribution

Machine learning has improved the potential for research in music pedagogy through its application to audio signal processing [7]. Systematic statistical analysis of sound features from the audio recordings can be done to assess the music performances automatically [8]. Of particular interest is the study of timbre quality in variable pitch instruments, such as bowed strings, woodwinds, and brasses, as it has an educational value and a connection to the performance quality [9–11].

Considering the trumpet as a case study, the role of timbre quality from both pedagogical and academic research perspectives has garnered considerable interest.

1.1 The musician's perspective of tone quality

Sound production on a trumpet involves a complex coordination and balance between the embouchure, the oral cavity, and the airflow [12]. Numerous literature sources in the field of music performance report how tone quality is closely related to the level of efficiency of the sound production mechanism of the player. It is a widespread belief in trumpet teaching that inefficiencies in playing caused by sub-optimal coordination of the muscular movements of the musician are reflected in the quality of the sound produced.

Arnold Jacobs, one of the foremost brass pedagogues of his time, reports how the presence of excessive muscle tension, which causes rigidity in the musician's body and inefficiency in playing, is reflected in a forced and strained sound [13]. Similarly, James Thompson argues that an incorrect and inefficient sound production mechanism obliges the player to force, leading to a decreased endurance and a strident tone [14]. Analogous observations are mentioned by Campbell et al. who argue that obtaining a good sound depends on the skill of the musician as well as the quality of the instrument, and that timbre plays a significant role in the ability to project the sound [15].

Similar conclusions are also drawn by Kristian Steenstrup, according to whom "it is more efficient, from the point of view of both the lip and respiratory musculature, to produce a beautiful, round tone rich in harmonics than a shrill or dull tone" and points to a lack of thorough research in this area [16]. Steenstrup proposes a parallelism between the sound production mechanism of trumpet players and that of singers, for whom there is more research in the literature. In particular, he hypothesizes that the influence between different types of phonation (i.e., pressed, flow, breath phonation) on the acoustic spectrum - as described by Sundberg and Gauffin [17] - may find a counterpart in the way sound is produced for brass musicians. It should be noted that this idea was proposed as a way to visualize or better understand sound production for trumpet players, without suggesting that the actual acoustic mechanisms involved are similar.

Other sources report that timbre provides most of the information to music teachers on which to base the choice of suggestions offered to their students, consequently influencing the determination of the educational path through targeted exercises [12, 18].

What is reported in the pedagogical field has also a correspondence in the context of orchestra auditions and competitions, where some of the rehearsals hide the performer from the jury's view in order to limit the introduction of bias in the selection of candidates. This suggests that an experienced ear, such as that of the jury members, is able to distinguish the level of efficiency and ease of sound production based solely on auditory information.

1.2 The researcher's perspective of tone quality

In the field of Music Information Retrieval (MIR), researchers are exploring the use of algorithms to identify the quality of trumpet tones from the audio information. The goal is to investigate the possibility of training a model that can assess the timbre quality of a trumpet sound like a qualified teacher. Knight et al. trained classifier models based on 56 mono and multidimensional audio features into two, three and seven classes [19]. Although the dataset used by this study, comprising only 239 notes played by four trumpeters, was limited, the results provided a promising proof of concept for future research in this field.

A few years later, the Music Technology Group at Pompeu Fabra University (MTG-UPF) conducted a research study on similar premises in collaboration with KORG Inc. targeting several musical instruments, including the trumpet [9, 20]. Having created an online platform for collecting, labeling and evaluating audio samples, they proposed a model for the assessment of the sound quality based on five attributes, namely dynamic stability, pitch stability, timbre stability, timbre richness, and attack clarity. Among these attributes, timbre richness - defined as the quality of timbre - is the one that comes closest to that considered by Knight et al. and suggested by brass teachers. The MTG-UPF study has significant limitations due to the lack of diversity in the dataset, consisting of sound samples collected from just two trumpet players. In order to train the model, the two musicians - professional music performance graduates - were required to record each of the attributes correctly and intentionally incorrectly. This could introduce further limitations in that it is not necessarily true that the sound of a professional musician who intentionally plays badly presents the same audio characteristics as the incorrect sound of a novice trumpet player, since the musician should be perfectly capable of dissociating from years of procedural memory training [21].

In this research, we aim to address the above limitations by training a model using an extensive dataset of sounds produced by 110 trumpet players of different expertise, to develop a system that has a better performance in real life applications. We report here our initial developments of this tool, as this is a work in progress. Finally, we present an interface for the application of this model in pedagogical contexts, in the framework of commercially available music education systems.

In Section 2, we describe the characteristics of the recorded dataset. Section 3 discusses the methodology adopted to develop the software and presents the results collected. Section 4 introduces the pedagogical framework of application. Finally, Section 5 discusses an interactive interface that implement the designed algorithm to support effective trumpet learning.

2. DATASET DESCRIPTION

Extensive data collection of trumpet sounds was conducted in music schools and masterclasses in low-noise environments with different acoustic conditions. Two sets of microphones were held 50 cm in front of the trumpet bell about 10 cm off its longitudinal axis for each recording.

The IM69D130 Shield2Go evaluation board manufactured by Infineon Technologies was used, equipped with two Infineon IM69D130 Micro-Electro-Mechanical Systems microphones, and interfaced to a Raspberry Pi Model 3B+ and a Raspberry Pi 4 Model B. Audio data were acquired with a bit depth of 32 and a sampling rate of 48kHz. These hardware components were selected since they use the same technologies as in mobile devices while ensuring a significantly high maximum sound pressure level (i.e., max SPL of 130dB) for recording the trumpet without distortion. This decision was made in order to consequently train a more robust model for use in mobile apps and software directly using the built-in microphone in smartphones, tablets, or PCs.

The complete dataset includes recordings of approximately 19,000 tones collected from 110 different trumpet players. The musicians included students from amateur music schools, students and teachers from conservatories and universities of the arts, music performance graduates, professional orchestral musicians and international soloists. The performers were playing their own trumpet with their own mouthpiece and were recorded using a set of two microphones held side by side. Tuning with respect to a reference pitch was not enforced, as the timbre quality is expected to be independent of reference pitches.

Each musician was asked to play separate long tones along a chromatic scale from E3 to B♭5 three times at different dynamics: once *piano*, once *mezzo forte*, and once *forte* in their order of preference. The duration of the notes, which is required to be at least 1 second in order to have sufficient margin to apply time-variant feature analysis afterwards, ranges from 0.7 to 4 seconds. Despite being asked to play separate tones, several beginners, for whom playing a chromatic scale in front of a recorder constituted a challenging task, played legato notes. Given their evident difficulty in managing the suggested tasks, and that the sounds acquired in that way still contained the timbral characteristics sought, they were not asked to repeat the recording.

To provide clearer guidance on the desired loudness variation (e.g., *piano*, *mezzo forte*, and *forte*), musicians were shown a digital sound level meter displaying the decibel level produced, along with the target reference levels (i.e., 85dB, 105dB, 115dB). This approach aimed to enhance the variability within the dataset, considering the significant impact of loudness level on timbre [22]. Although the performers did not always maintain the exact dB level requested (particularly beginners), the presence of the sound level meter proved beneficial in directing their attention towards differentiating the loudness levels they were playing.

Although musicians were asked to play a chromatic scale along the entire frequency range of the trumpet (i.e., E3 to Bb5), the recordings of trumpet players with less expertise is not complete, missing the notes in the highest register. Indeed, it is common knowledge that playing high notes with the trumpet requires an advanced level of control and coordination between embouchure, oral cavity, and air flow.

The audio dataset described above was independently collected by the first author before the start of his academic program at the host institution. The recording conditions, player's level of expertise and an overall grade on the quality of the sound produced were annotated at the time of recording each track.

The collected dataset encompasses a diverse array of audio samples, providing a suitable environment for training a machine learning model to differentiate the timbre quality level of a trumpeter in real-world scenarios. Section 3 outlines the methodology applied to tackle this challenge, resulting in the development of a robust model that can accurately assess the tone quality of trumpet performances.

3. METHODOLOGY (ML MODEL)

In this section, we present the preliminary work on the analysis of the collected dataset and training a machine learning model to distinguish the efficiency of a trumpet player's sound production from the audio data.

The audio dataset described in the previous section was segmented into single trumpet tones using the pyin vamp plugin by Mauch and Dixon [23]. While the segmentation was accurate in most cases, some issues were noted upon closer inspection. The issues included clipped audios, background talk being segmented as a note, etc. It was thus deemed necessary to manually listen to the notes before training a machine learning model directly using the preliminary grades as the labels. Due to practical difficulty in listening to each of the 19,000+ notes, we decided to base the machine learning model on a sufficiently large and balanced subset of the complete dataset.

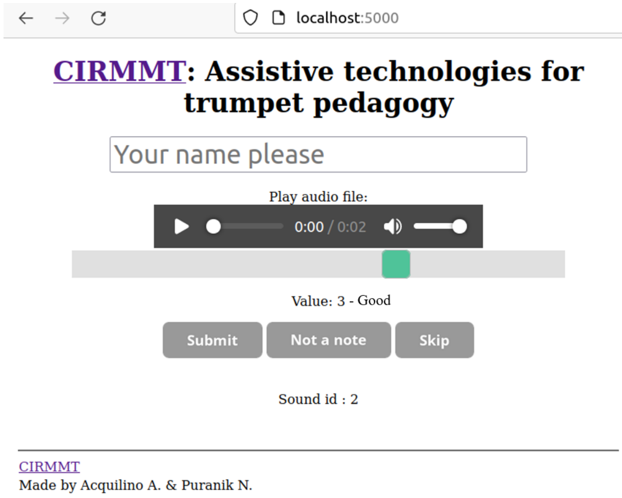


Figure 1. Interface for blind grading the trumpet notes

In this study, we aimed to develop a 4-level quality classifier for the sounds in the audio dataset, as proposed by Wesolowski [24], aiming to simplify the label assignment process while maintaining sufficient variability. The classifier levels were defined as 1:poor, 2:fair, 3:good, 4:excellent. From the preliminary grades assigned at the time of recording, seventeen players were selected: five of them were assigned to one class and four performers to each of the other classes. Indeed, as beginners were unable to perform the complete chromatic scale, one player more was assigned to the 1:poor category in order to counterbalance the reduced representation of audio samples within that particular class. This selection gave a total of 1,711 single notes which were manually labeled by the first author who is a graduate in trumpet performance with professional experience as a musician and a brass teacher.

To facilitate the labeling process, an online interface was developed for assigning unbiased grades. The single notes of the selected musicians were uploaded to the web page, where the grader could listen repeatedly without knowing the identity of the musician or their class. The grader could then assign a grade from 1 to 4, corresponding to the level of timbre quality, as shown in the Figure 1.

In the case where a sound had been poorly segmented, the grader could indicate that the sound was not a note. The decision to evaluate one tone at a time was made to identify and remove wrongly segmented notes, thus obtaining a clean dataset of 1,481 tones. On analysis of the ratings assigned to individual notes through the web interface and the ratings assigned to the full performance of the musicians at the time of recording, we observed that they had a Pearson correlation coefficient of 87.5% (p-Value$<$0.001). The high correlation may not appear surprising given that the same person (the first author)

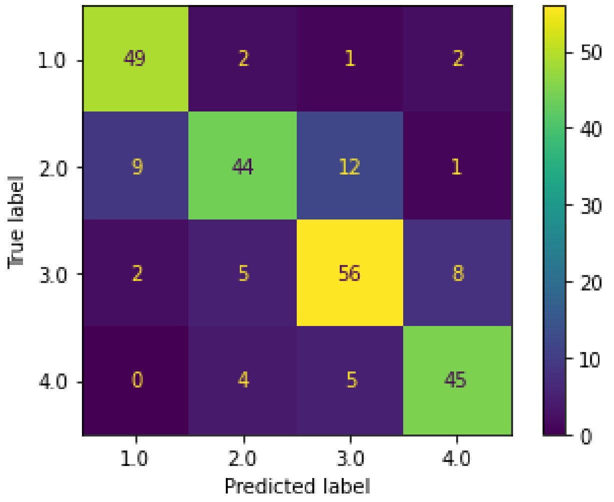


Figure 2. Confusion matrix for a model trained with 75-25 train-test split

made both the ratings. However, it is significant that the first author's perception of the tone quality at the aggregate and the granular levels remained consistent.

The cleaned audio dataset of 1,481 tones was used to fit a Random Forest Classifier model [25], an ensemble learning method that leverages multiple decision trees for enhanced accuracy in classification tasks. As a preprocessing step, the sound samples were first scaled to have a maximum signal amplitude equal to one and white noise at -60dB was added to the audio. The audio features for each tone were then extracted using the essentia Extractor algorithm [26]. To reduce the computational complexity, we used only the statistical aggregates of the audio features such as mean, variance, mean of derivative, etc. As a first step, we included all except the rhythm based features (a total of 1,230 features) to fit a Random Forest Classifier. For a 10-fold cross-validation we achieved a mean accuracy score of 78%.

We used the Random Forest Classifier model-based feature selection to identify the top 256 features for the classification. Using only these features, the 10-fold cross validation mean accuracy score improved slightly to 81.37%.

We fit a model by splitting the dataset into two parts: 75% of the data is used for training (i.e., train set) and the remaining 25% is used for testing (i.e., test set). The confusion matrix, which indicates the performance assessment of the model on the test set, can be seen in Figure 2. It can be observed that the most confusion is between the adjacent classes. Since the audio samples in the adjacent classes are in fact more similar to each other than the other classes, the errors seem to be reasonable. In fact, it is quite possible that for some audio samples there would be such a disagreement in the opinions of two human experts, or the same expert at different times.

To evaluate the out-of-dataset performance of the trained model, the latter was tested using the sound samples of trumpet from the Good-sounds dataset curated by Pompeu Fabra University [9,20], as previously described in Section 1.2. The Good-sounds dataset has 25 notes labelled as 'bad timbre richness'; however, there is no specific label for the 'good timbre richness' notes. Therefore, the model was tested on sounds labelled as 'good', which may include different tonal qualities (e.g., pitch stability, dynamics stability) in addition to 'timbre richness'. All notes labeled as 'bad timbre richness' were given grade 1, while about 70% of the notes labeled as 'good' received a grade 3 or 4 by the model. The results are reported in Table 1.

Good-sounds labels	**Model grades**			
	1	**2**	**3**	**4**
Bad (n=25)	100%	0%	0%	0%
Good (n=190)	30%	0.5%	9.5%	60%

Table 1. Effectiveness of trained RF Classifier Model tested on the Good-sounds database

An analysis of the sounds labeled as 'good' to which a grade equal to 1 was assigned reveals that those notes predominantly had a high vibrato extent, significantly more than was present in the training dataset.

These findings were further substantiated by conducting informal trials, wherein the model was assessed in real time alongside graduate trumpet players, thereby suggesting promising potential for pedagogic applications. Subsequent training with a set of experienced trumpet teachers is in progress.

4. PEDAGOGICAL FRAMEWORK

The field of edTech systems for assessing trumpet tone quality is relatively limited. Furthermore, the few technologies that address this issue may be limited in their pedagogical effectiveness, as discussed by Acquilino and Scavone [27]. One example of this is KORG cortosia[1], which estimates the trumpet sound "goodness" by rating in real time five elements: pitch stability, dynamic stability, timbre stability, timbre richness, and attack clarity. An overall numerical score is provided along with a five-axis visualization corresponding to each of the five elements. The complexity of the designed interface may have affected perceived ease of use and limited its adoption by music students.

Other systems, such as TonalEnergy[2], provide real-time visualization of the spectrum and/or height of partials. However, this kind of feedback is likely too complex to be interpreted by music students and even teachers, given the high number of variables involved in identifying tone quality.

This suggests the need for further development of user-friendly and pedagogically effective edTech solutions for assessing trumpet tone quality. A call for new technologies to improve self-regulated music learning, other than student-teacher interaction, is also supported by a recent thorough study conducted by Waddell and Williamon [28].

[1] www.korg.com/us/products/software/cortosia
[2] www.tonalenergy.com

Where effective musical instrument learning is concerned, the deliberate practice framework introduced by Ericsson et al. [29] assumes a crucial role, given its acknowledged success in research and music performance. Deliberate practice refers to the use of specific and targeted activities, with immediate feedback, that are designed to improve specific aspects of performance. Specifically, Ericsson and Harwell [30] expressed the following four criteria that define purposeful/deliberate practice:

1. *The practice involves individualized training of a trainee by a well-qualified teacher. This teacher can assess which aspects a particular trainee would be able to improve during the time until the next meeting and is able to recommend practice techniques with established effectiveness.*

2. *The teacher must be able to communicate the goal to be achieved by the trainee and that the trainee can internally represent this goal during practice.*

3. *The teacher can describe a practice activity to attain the identified goal for performance and that this activity allows the trainee to get immediate feedback on a given attempt.*

4. *The trainee is able to make repeated revised attempts that gradually approach the desired goal performance.*

Criterion 1 emphasizes the significance of an experienced teacher in facilitating the learning of music students. This suggests the idea of developing technologies that support brass instructors in becoming more effective, extending their pedagogical potential. In addition, Criteria 2, 3 and 4 provide guidelines for the design of such technologies, particularly with regard to the role they can play in assisting the adoption and engagement of deliberate practice. Criterion 2 suggests introducing new ways for music students to understand the concepts of technical skills to be learned with their instruments. This could be achieved through a multi-sensory experience (e.g., combining auditory and visual feedback). Criterion 3 recommends the design of technologies that provide real-time feedback to musicians. Criterion 4 highlights significant pedagogical potential toward systems that can objectively quantify the level of accuracy of sound produced by the performer with reference to a specific technical skill (in this case, timbre quality). In fact, specific synchronous feedback facilitates students' growth in achieving the targets set with the teacher [31]. In this way, music students can objectively track their progress over time and identify strengths and weaknesses for targeted practice.

The assumption is that designing educational systems that follow the criteria above will result in increased perceived usefulness among teachers and music students. The Technology Acceptance Model by Davis explains that these perceptions of usefulness and ease of use create an intention to use the technology, and this intention leads to actual usage [32].

Chapter 5 presents a technological exercise that aims to fulfill the premises discussed in order to stimulate effective study of tone quality by trumpet students.

5. PROPOSED EDUCATIONAL TECHNOLOGY

This section presents an innovative musical exercise aimed at promoting the acquisition of timbre qualities by trumpet students. The exercise is designed to be interactive in real-time, allowing for flexible adjustments by the teacher.

The proposed edTech system prompts the user to enter the metronome value, the lowest and highest note they wish to play between E3 and B♭5. These values are fixed and shown in the top row of the interface in Figure 3. Thus, the exercise starts by presenting a musical score of three four-quarter measures with a random note to be played between the minimum and maximum values selected, as shown in the third row of Figure 3. The first measure is a pause, the second measure shows in gray the selected note that the software plays in time with a trumpet sound of high timbre quality, and the third measure shows the same note to be played by the musician. At the end of the third measure, another score is generated with another randomly selected note and the sequence is repeated for a given number of times. In order to aid proper time keeping, a metronomic "tic" and a synchronized numerical display of the corresponding beat number in the first column of the second row of Figure 3 are included. The sound played by the software during the second measure was previously recorded by a professional trumpet player. It was chosen to use a recording of a fine performer, instead of simply using MIDI sounds, to induce the user to emulate a high-quality timbre by imitation. In order to broaden the range of stimuli for students, multiple high-quality recordings can be provided for each note and randomly selected.

The model presented in Section 3 is implemented within the software by providing near real-time feedback on the level of tone quality played by the user. The sound is analyzed by the algorithm which returns feedback corresponding to the class in which it is associated. The four different emoticons 🤔 😊 😃 🤩 are associated with classes 1, 2, 3, 4 respectively appearing in the second row on the right side of the interface in Figure 3.

This exercise aims to stimulate deliberate practice in music students regarding the acquisition of high-quality timbre, which many brass pedagogues associate with an efficient sound production mechanism. The inclusion of the model within an interactive exercise with gamified experience is intended to foster users' motivation. The possibility of defining the range of notes that can appear in the score ensures that the teacher can gradually approach trumpet students toward the high register. In fact, in the book *Brass Techniques and Pedagogy*, Weidner reports as a common problem among trumpet players the change in tone quality when playing in the high register during lip slurs [33]. The author states that it can be counterproductive to try to play high notes with the trumpet without being able to play with good tone quality lower notes, pointing out as potential undesirable effects the development of bad habits involving excessive

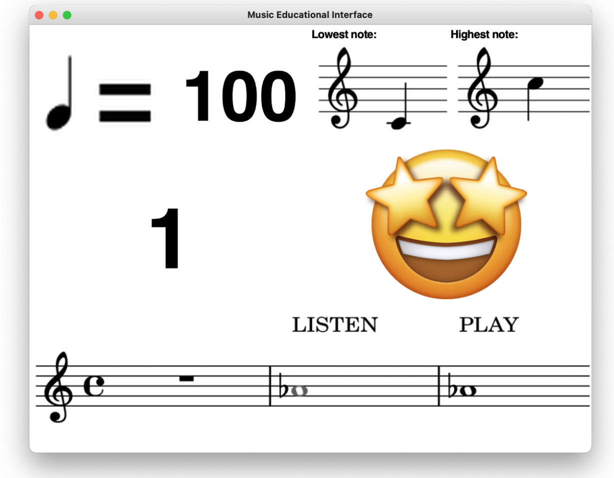


Figure 3. Interface of the proposed edTech system. Top row from left to right shows the input parameters: metronome value, lowest note, and highest note. The middle row shows the current beat value on the left and an emoji feedback on the quality of the timbre produced on the right. At the bottom, a three-measure score is shown cyclically updating with a random note to be played between the highest and lowest values selected. Additional resources about the interface can be found at `https://pninad.github.io/smac2023/`

pressure, lip pinching, and poor air support. The exercise presented thus offers the teacher control over the students' note range development while also providing direct real-time feedback on the quality of tone played. This is intended to make the user focus on the development of this skill by stimulating deliberate practice while avoiding the aforementioned problems.

The developed interface could also provide music instructors with a summary of students' achievements between in-person lessons. In this way, they can observe their learning curve in acquiring sound efficiency, identify learning criticalities, and propose targeted practices to overcome them. The algorithm devised can also be flexibly applied to other interfaces. The mere functionality of emoticons that change according to the level of timbre quality can, for example, be used directly in stand-alone mode, such as chromatic tuners. Or, trumpet teachers can enter the sequence of notes for students to play, instead of relying on a random choice.

6. CONCLUSIONS

This paper presented the topic of timbre quality for the trumpet from both the perspective of music performance teachers and scientists. Through the use of an extensive dataset, the researchers were able to overcome the limitations of previous work and create a model that can distinguish the level of tone quality with a mean 10-fold cross validation accuracy of about 80%.

The authors also proposed an innovative edTech system that utilizes the Technology Acceptance Model framework to encourage deliberate practice in trumpet students, with the goal of improving their tone quality. The proposed interactive exercise can be flexibly adapted by the teacher to meet the individual needs of their students. This approach has the potential to not only enhance trumpet timbre quality, but also to improve teacher-student interaction.

A limitation of the present investigation concerns the use of only a single rater for the training of the model. Such an approach may be susceptible to the influence of rater bias, consequently restricting the generalisability of the results. In an effort to mitigate this inherent limitation, the authors are undertaking a study that incorporates multiple raters in the process of assigning labels to the dataset. This approach is intended to facilitate the training of a model that is less susceptible to the impact of potential individual biases, thereby enhancing the robustness and applicability of the findings.

Future research aims to expand this study exploring the acoustical properties of trumpet sound efficiency by analyzing the main discriminatory audio features identified by the model. In addition, a further investigation is ongoing to evaluate the effectiveness of the proposed system in music pedagogy, focusing on its ability to support trumpet students in learning sound production efficiency.

Overall, this work aims to provide novel solutions in the field of technology-enhanced music learning and provides a valuable resource for trumpet teachers, students, and researchers alike.

Acknowledgments

This work was made possible with the support of a CIRMMT Student Award and a Tomlinson Doctoral Fellowship.

The authors thank the foundational contribution of Mirko d'Andrea and Emanuela Bussino for the audio data collection stages, as well as all the volunteers and colleagues who supported this research.

7. REFERENCES

[1] S. Raptis, A. Chalamandaris, A. Baxevanis, A. Askenfelt, E. Schoonderwaldt, K. F. Hansen, D. Fober, S. Letz, and Y. Orlarey, "Imutus - an effective practicing environment for music tuition," in *Proceedings of 2005 International Computer Music Conference*, Barcelona, Spain, 2005, pp. 383–386.

[2] S. Bagga, B. Maurer, T. Miller, L. Quinlan, L. Silvestri, D. Wells, R. Winqvist, M. Zolotas, and Y. Demiris, "Instrumentor: An interactive robot for musical instrument tutoring," in *20th Annual Conference on Towards Autonomous Robotic Systems*, 2019, pp. 303–315.

[3] Y. Tsuji and N. Atsuhiro, "Development and evaluation of drum learning support system based on rhythm and drumming form," in *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, 2006, pp. 11–21.

[4] E. Schoonderwaldt, A. Askenfeld, and K. Hansen, "Design and implementation of automatic evaluation of recorder performance in imutus," in *Proceedings of 2005 International Computer Music Conference*, International Computer Music Association, Barcelona, Spain, 2005, pp. 97–103.

[5] M. Robine, G. K. Percival, and M. Lagrange, "Analysis of saxophone performance for computer-assisted tutoring," in *Proceedings of the 2007 International Computer Music Conference*, 2007, pp. 381–384.

[6] G. K. Percival, Y. Wang, and G. Tzanetakis, "Effective use of multimedia for computer-assisted musical instrument tutoring," in *Proceedings of the International Workshop on Educational Multimedia and Multimedia Education*, 2007, pp. 67–76.

[7] A. Lerch and P. Knees, "Machine learning applied to music/audio signal processing," in *Electronics*, 2021.

[8] P. Seshadri and A. Lerch, "Improving music performance assessment with contrastive learning," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2021, pp. 634–641.

[9] G. Bandiera, O. Romani, H. Tokuda, W. Hariya, K. Oishi, and X. Serra, "Good-sounds.org: a framework to explore goodness in instrumental sounds," in *Conference: 17th International Society for Music Information Retrieval Conference*, New York, 2016.

[10] S. Giraldo, R. Ramirez, G. Waddell, and A. Williamon, "A computational approach for measuring performance quality in violin tones," in *International Symposium in Performance Science*, Reykjavik, 2017, pp. 2231–2240.

[11] S. Giraldo, G. Waddell, I. Nou, A. Ortega, O. Mayor, A. Perez, A. Williamon, and R. Ramirez, "Automatic assessment of tone quality in violin music performance," in *Frontiers in Psychology*, 2019, pp. 1–12.

[12] F. G. Campos, *Trumpet technique*. Oxford University Press, 2005.

[13] A. Jacobs and B. Nelson, *Also Sprach Arnold Jacobs: A Developmental Guide for Brass Wind Musicians*. Polymnia Press, 2006.

[14] J. Thompson, *The buzzing book complete method; trumpet or other brass instruments*. Editions BIM, 2003.

[15] M. Campbell, M. Gilbert, and A. Myers, *The Science of Brass Instruments*. Modern Acoustics and Signal Processing, 2021.

[16] K. Steenstrup, *Teaching Brass*. Det Jyske Musikkonservatorium, 2007.

[17] J. Sundberg and J. Gauffin, "Waveform and spectrum of of the glottal voice source," in *Frontiers of speech communication research*, London: Academic Press, 1978, pp. 301–320.

[18] G. Cassone, *The trumpet book*. Zecchini, 2009.

[19] T. Knight, T. Upham, and I. Fujinaga, "The potential for automatic assessment of trumpet tone quality," in *Proc. of the 12th Int. Society for Music Information Retrieval Conf., ISMIR*, 2011, pp. 573–578.

[20] O. Romani, H. Parra, D. Dabiri, H. Tokuda, W. Hariya, K. Oishi, and X. Serra, "A real-time system for measuring sound goodness in instrumental sounds," in *138th Audio Engineering Society Convention, AES*, Warsaw, Poland, 2015, pp. 1106–1111.

[21] M. Lam, "The physicality of music production: Investigating the roles of mindful practice and kinesthetic learning," in *Music Educators Journal*, 2020, pp. 23–28.

[22] D. Luce and M. J. Clark, "Physical correlates of brass-instrument tones," in *The Journal of the Acoustical Society of America*, 1967, pp. 1232–1243.

[23] M. Mauch and S. Dixon, "Pyin: A fundamental frequency estimator using probabilistic threshold distributions," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 659–663.

[24] B. C. Wesolowski, "Understanding and developing rubrics for music performance assessment," in *Music Educators Journal*, 2012, pp. 36–42.

[25] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1. IEEE, 1995, pp. 278–282.

[26] D. Bogdanov, N. Wack, G. E. E., S. Gulati, P. Herrera, and O. M. et al., "Essentia: an audio analysis library for music information retrieval," in *International Society for Music Information Retrieval Conference*, 2013, pp. 493–498.

[27] A. Acquilino and G. Scavone, "Current state and future directions of technologies for music instrument pedagogy," in *Frontiers in Psychology*, 2022.

[28] G. Waddell and A. Williamon, "Technology use and attitudes in music learning," in *Frontiers in ICT*, 2019.

[29] K. A. Ericsson, R. T. Krampe, and C. Tesch-Römer, "The role of deliberate practice in the acquisition of expert performance," in *Psychological Review*, 1993, pp. 363–406.

[30] K. A. Ericsson and K. W. Harwell, "Deliberate practice and proposed limits on the effects of practice on the acquisition of expert performance: Why the original definition matters and recommendations for future research," in *Frontiers in Psychology*, 2019.

[31] L. Martin, “Using synchronous formative feedback to facilitate student growth,” in *Music Educators Journal*, 2020, pp. 51–57.

[32] F. D. Davis, “Perceived usefulness, perceived ease of use, and user acceptance of information technology,” in *MIS Quarterly*, 1989, pp. 319–340.

[33] B. Weidner, *Brass Techniques and Pedagogy*. Palni Press, 2020.

Transversal bending is an essential aspect of saxophone reed dynamics

Enis Ukshini, Joris Dirckx
Department of Physics, Laboratory of Biomedical Physics BIMEF, University of Antwerp, Groenenborgerlaan 171, Antwerp 2020, Belgium
enis.ukshini@uantwerpen.be
joris.dirckx@uantwerpen.be

ABSTRACT

Stroboscopic digital image correlation allows to measure saxophone reed displacement and strain over the entire surface of the reed which is not covered by the lip. During the vibration cycle, at the moment where the reed touches the tip and side rails of the mouthpiece, the reed bends along the direction perpendicular to its long axis. Due to this bending, the center part of the reed tip is lower than its circumference, and compressive strain suddenly develops. This paper investigates the bending of a reed on a normal mouthpiece and a mouthpiece with a thin supporting ridge placed underneath the reed and parallel to the reed fibers so that it has minimal influence on the airstream. No vibration is obtained when the ridge is as high as the side rails. With a ridge 0.2 mm lower than the side rails, bending and strain were similar to the normal mouthpiece. However, an extra bounce effect of the reed on the mouthpiece was observed in the displacement curve. A professional player evaluated both mouthpieces. The high ridge mouthpiece was found to be unplayable. The mouthpiece with a slightly lower ridge influenced tone quality but was found to be playable. In conclusion, the transversal bending of the reed is an essential aspect of reed dynamics, and bidirectional elasticity parameters are important.

1. INTRODUCTION

In woodwind instruments with single reeds, little research has been conducted on reed mechanics and its impact on sound quality and playability [1]. The vibrations of the saxophone reed are determined by a complex interplay between the mechanical parameters of the reed, the air pressure above and beneath the reed surface, and the acoustic feedback of the instrument. In attempts to model reed vibration, only the elastic parameters along the length axis have been considered [2], [3]. Recent work has shown that stroboscopic digital image correlation (DIC) makes it possible to measure deformation on the whole freely vibrating part of the reed, which is not covered by the lip [4]. DIC also allows to determine full field strain maps on this surface.

While longitudinal bending is the main part of reed motion, the transversal bending of the reed has been largely neglected in previous studies [5]–[9].

This research paper investigates the transversal bending of the reed in three different mouthpiece designs. The bending of the reed is analyzed in each mouthpiece using constant phase stroboscopic digital image correlation, which will allow to visualize and quantify the transversal bending motion of the reed. A modified 3D printed mouthpiece is used to investigate the effect of transversal bending.

2. METHODS

2.1 Reed and mouthpiece

To avoid dehydration artifacts, tests were performed using high-quality synthetic reeds (Légère Signature Series, Strength 2.75). For DIC, the studied object surface needs to have an optical texture, preferably consisting of random speckles with high contrast. As the reeds are semitransparent, a thin layer of white paint was first applied. Then, an airbrush was used to apply a very fine pattern of random black speckles. To obtain high contrast without specular reflections, paint with an absorbance of more than 99 % (VantaBlack, Musou Black) was used for the speckle pattern.

Using X-ray microtomography and high-resolution 3D printing, a replica was made of a professional high quality classical mouthpiece (Concept, Henri Selmer, Paris). Both tomography and printing had a resolution of better than 50 micrometers. The tablet and side rails of the mouthpiece were further post-processed using grain 2000 sandpaper to obtain a perfectly flat finish. Using 3D design software, two altered replicas were produced, each with a thin vertical ridge placed on the centerline of the baffle. Figure 1a shows a 3D rendering of the altered mouthpiece; figure 1b shows a cross-section at 7 mm from the mouthpiece tip. In the first altered mouthpiece, the height of the ridge was equal to the height of the mouthpiece side rails. In the second version, the height was 0.2 mm lower than the side rails.

The original Concept, its replica, and the two altered replicas were used to measure reed vibration and reed strain with the DIC setup on note A3 (alto saxophone notation). In a previous paper, the range of blowing pressures and lip forces was investigated for which reed vibration occurs [10]. For the present measurements, average values of 4 kPa blowing pressure and 2 N lip force will be used for A3.

a)

b)

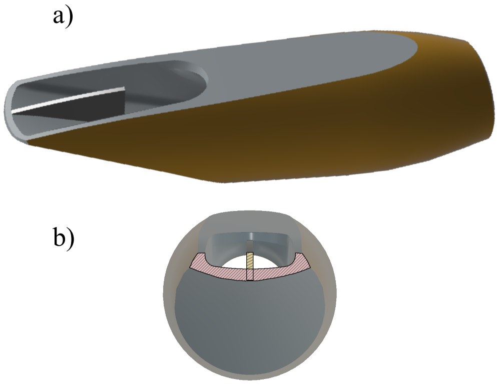


Figure 1: (a) 3D rendering of an altered Concept mouthpiece design. A thin vertical ridge is placed on the centerline of the baffle. (b) cross-section at 7 mm from the mouthpiece tip.

Figure 2a shows the mouthpiece with the reed mounted. Figure 2b shows a schematic sagittal section of the mouthpiece. Some names of the mouthpiece parts are indicated for further reference (table, side rails, tip), and the coordinate system that will be used to present the results is also indicated. The lip leaves a reed length of about 10 mm, free to vibrate. This is the visually accessible part where measurements can be performed. The Y-axis is chosen parallel to the reed fibers; the Z-axis is determined perpendicular to a plane fitted through the reed after lip force has been applied. This direction is close to the normal on the mouthpiece surface, but it makes it easier to report reed displacement values correctly. The direction of the Z-axis is chosen towards the mouthpiece so that displacements of the reed towards the mouthpiece have positive values.

The reed rests on the top surface of the mouthpiece, called the table, and partly on the side rails. The table is flat, but the side rails bend away towards the mouthpiece's tip so that a slit gradually develops between the two. After mounting, a narrow slit, typically between 1 and 2 mm wide, remains between the reed's front tip and the mouthpiece's front tip. The musician’s lower lip presses the reed toward the mouthpiece while playing.

Once lip force is applied, blowing pressure further pushes the reed toward the mouthpiece. As pressure increases, the width of the slit decreases, thus diminishing flow. This negative hydrodynamic resistance is the source of the oscillation energy. When the reed is pushed close enough, oscillation starts. The acoustic properties of the instrument determine the oscillation frequency [11], [12].

The reed position with lip force applied (but without blowing pressure) will be used as the reference to report on the dynamic displacements.

2.2 Measurement setup

In the measurement setup, the mouthpiece sits in an airtight transparent box, and the neck passes through the top lid of the box with an airtight seal. Figure 3 shows a schematic diagram of the measurement setup. The side view (3a) shows how the neck is mounted in the box. The lip can be moved towards the mouthpiece using a stepper motor actuated translation stage, and lip force is measured with a load cell mounted on this stage. The instrument itself is in ambient air. A ventilator system was used to generate an overpressure in the box, to imitate the blowing by the musician [13]. An artificial lip made of silicone rubber was pushed against the top surface of the reed, leaving a length of about 10 mm free to vibrate. The position and shape of this lip were determined in a preparatory experiment where lip prints were taken from a set of players.

a)

b)

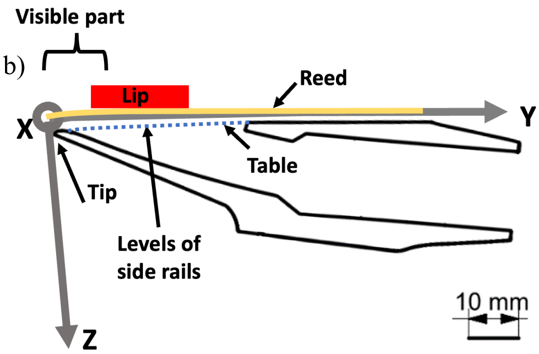


Figure 2: (a) Saxophone mouthpiece on a saxophone neck, with the reed mounted on the mouthpiece. (b) Schematic drawing of a sagittal section of a mouthpiece, with the coordinate system in which the data will be presented. The origin of the x-axis is set at the middle of the mouthpiece tip. The z-axis is chosen perpendicular to the surface of the reed when lip force applied. Measurement data will be presented on the visible part of the reed, where it is not covered by the (artificial) lip.

The top view on the setup (figure 3b) shows the positioning of the DIC cameras relative to the mouthpiece. The cameras observe the reed's free part through the transparent box's front wall. Further details on the measurement setup can be found in a previous paper [14].

Two CCD cameras (Manta G609B, Allied Vision) were used to perform the DIC measurements, placed under an angle of 13.5° to the left and right of the Z-axis. Only the part of the reed not covered by the lip can be measured. The cameras have 2752 by 2206 pixels resolution, and the reed is 16 mm wide.

The reed was illuminated using a high-power 1W LED placed outside the artificial mouth. Stroboscopic illumination with a duty cycle of 1/50th of the vibration period was used. A custom-made triggering system synchronized the stroboscope to the reed vibration. The length of the stroboscopic pulses is 2% of the vibration period. The exposure time of the camera is 200 ms, so that at a vibration frequency of 262 Hz, 52 stroboscopic pulses are integrated on the camera target to obtain sufficient illumination. A microphone picked up the sound generated by the saxophone. From this sound signal, a custom-made low-pass and zero detection circuit generated the trigger signal for

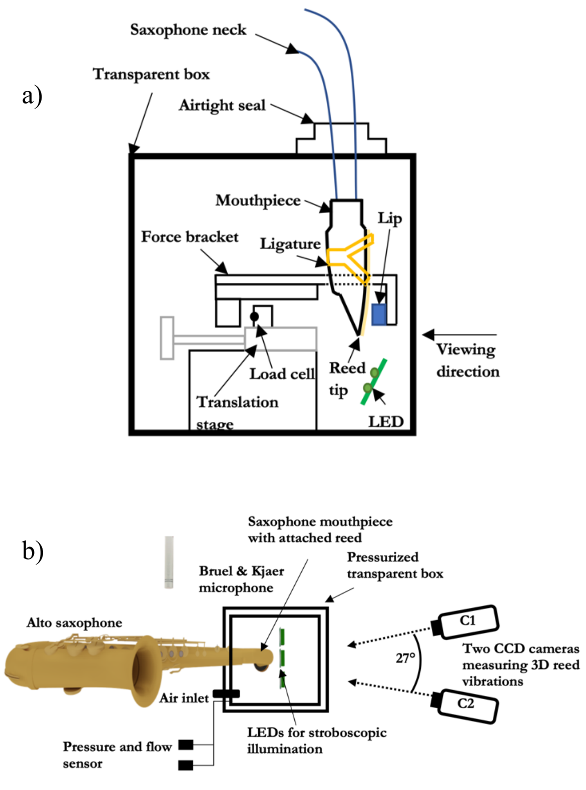


Figure 3: Schematic diagram of the measurement setup. (a) The side view shows how the neck is mounted in the artificial mouth. (b) Top view. Two cameras (C1, C2) observe the front of the reed.

the stroboscope. For each measurement, a set of 50 images was acquired. Between each image, the phase of the stroboscope was shifted over 1/50th of the vibration period. The phase shift was controlled by the frame_readout trigger signal of the camera so that images with the correct stroboscope phase could be acquired at the highest possible speed without the need for PC software to change the stroboscope phase. An entire vibration phase could be measured in less than 12 seconds.

DIC calculations were performed using Dantec ISTRA4D software. The software determines the measurement precision by an error propagation technique and considers all parameters (setup, calibration, speckle pattern quality). For out-of-plane precision, a measurement precision of less than 1 µm was found. X and y strain could be measured with a precision of 0.01 %. To evaluate the DIC results, the subset and step sizes were set to 39 and 20 pixels, respectively. A grid reduction factor of 2 and a spline filter has been used to smooth the results.

2.3 Subjective evaluation

The mouthpieces were given to a professional classical saxophone player (>30 years of experience) who had been using the Concept mouthpiece daily for several years. The player was instructed to test the mouthpieces for playability and sound quality. The player was instructed not to observe the shape of the mouthpiece. Each mouthpiece was identified with an inscription, which was only known to the investigator. The player used a Likert scale [15] to describe the playing experience, with the possibility to add free comments. The scale varied from 0 to 10, with 0 meaning "far worse than Concept," 5 meaning "equal to Concept," and 10 meaning "far better than Concept," and the scale was used to judge "playability" and "sound quality."

3. RESULTS

3.1 Objective measurements

3.1.1 Original Concept and replica

First, measurements were performed on the original Concept mouthpiece. Figure 4a shows a full field deformation map of the visible part of the reed at the moment in the vibration period where the reed touches the mouthpiece's tip. Figure 4b shows the strain distribution in the x-direction, and figure 4c shows the strain distribution in the y-direction at the same moment.

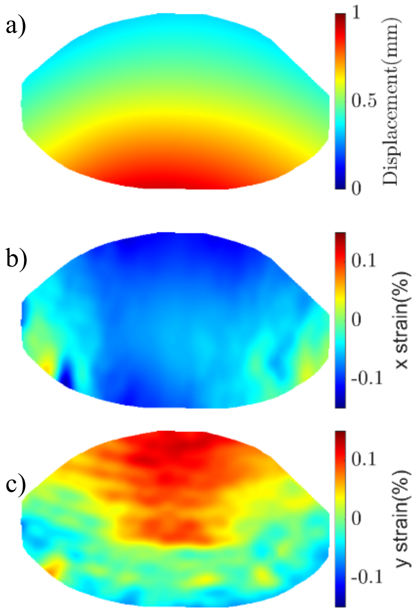


Figure 4: **Original Concept** a) displacement map measured at the moment when the reed touches the mouthpiece. Simple bending along the y-axis is mainly seen. b) x-strain map at the same moment. X-strain is mainly compressive. c) y-strain map at the same moment. Strain is mainly expansive.

At first glance, the deformation map only shows bending along the y-direction, which is the movement's main component. Some small dependencies on the x-position can be seen. The strain maps reveal the effect of bending along the x-direction more clearly: in the center zone of the measured region, a clear compressive strain exists. The absolute values of both x and y strains are small (0.1% and 0.2%, respectively). However, the sensitive DIC method allows us to measure them.

Figure 5a shows displacement as function of time for a single point in the center of the visible part of the reed. The colored dots indicate the moments within the vibration phase where measurements were taken. Figures 5b and c show displacement data in a direction parallel to the x-axis for the same time points. The level where the section was taken is indicated in the inset in figure 5b.

A slight asymmetry is seen in the cross-sections due to imperfect lip positioning. The figure clearly shows that the reed is bending along the x-direction: its center point is 0.15 mm below the level of the side rails.

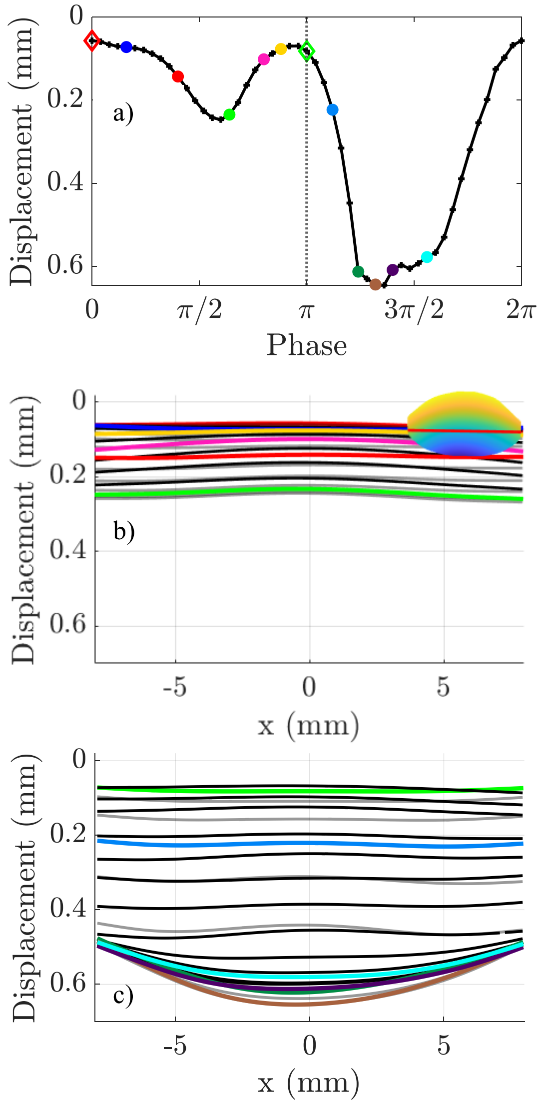


Figure 5: **Original Concept** (a) displacement as a function of time of a point across the cross-section indicated in (b). The black dots indicate the 50 equidistant measurement points. The red and green diamond respectively indicate the start of the first and second half of the vibration phase. The colored dots indicate moments in the vibration phase on which data will be presented in more detail. (b) displacement along a line parallel to the x-axis as a function of time for the first half of the vibration cycle. Location of the cross-section is shown in the inset. (c) displacement as a function of time for the second half of the vibration cycle for x cross-sections.

Figure 6 shows the x-strain and y-strain along the same cross-section for the colored time dots. At both extremes of the cross section, y strain is compressive. In the middle part of the reed, y strain is expansive and most prominent. The highest measured value was 0.13 %. This value was measured when the reed started moving away from the mouthpiece (around $3\pi/2$) after touching the mouthpiece and not when the reed touched the mouthpiece.

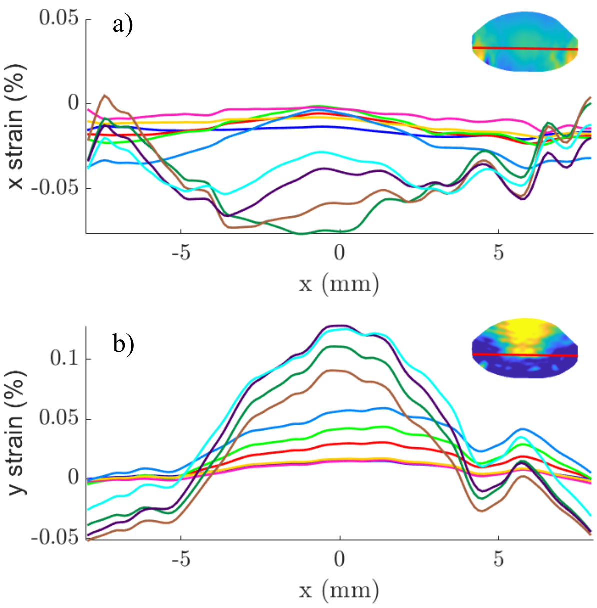


Figure 6: **Strain along a section parallel to the x-axis for the original Concept mouthpiece** (a) x-strain at the moment in time indicated in figure 5. (b) y-strain for the same time points. Y-strain increases the most at the middle of the reed when the reed moves towards the mouthpiece.

X-strain is compressive along nearly the entire section. The highest compressive value (0.08 %) was observed in the middle of the reed. When the reed is the furthest away from the mouthpiece, x-strain is nearly zero (<0.01 %). The highest x-strain value was measured for the moment in time before the reed touches the mouthpiece.

For the replica, very similar data were obtained. As an example, figure 7 shows the deformation and strain maps. The deformation is nearly identical to the data obtained on the original mouthpiece. The strain curves are also very similar; subtle differences are due to measurement noise and minor differences in reed and mouthpiece mounting and lip(force) adjustments.

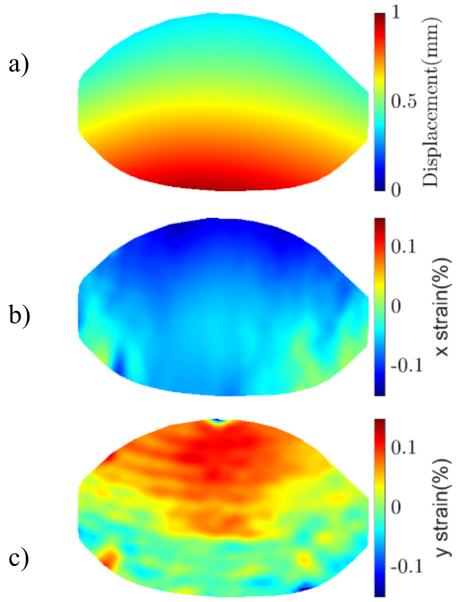


Figure 7: **Concept replica** a) displacement map measured at the time instant the reed touches the mouthpiece. Simple bending along the y-axis is mainly seen. b) x-strain map at the same time instant. X-strain is mainly compressive. c) y-strain map at the same moment. Y-strain is mainly expansive. The results are similar to the original Concept mouthpiece.

3.1.2 Altered replicas

For the replica with the high ridge, no data could be obtained. With the chosen lip force setting and blowing pressure, it was impossible to set the reed into vibration. For the replica with the low ridge, it was perfectly possible to set the reed into vibration. Data of full field deformation, x and y strain are shown in figure 8. When comparing the three data types, the x-strain data was the most prominent in observing differences between the three mouthpiece designs. Generally, the full field map of x-strain for the mouthpiece with the low ridge showed less compression at the middle of the measured reed surface.

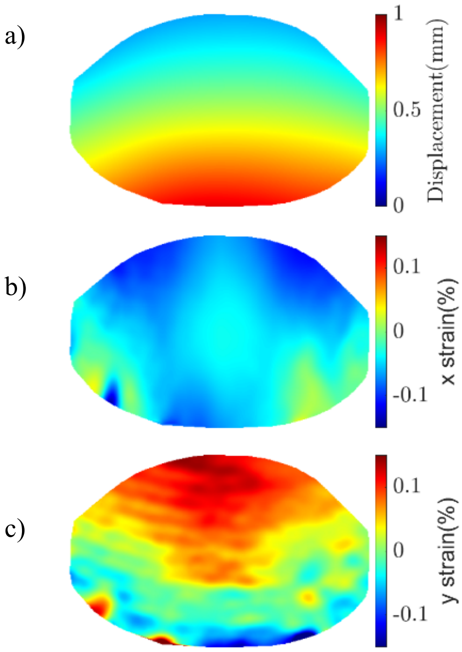


Figure 8: **Concept low ridge** a) displacement map measured at the moment when the reed touches the mouthpiece. Simple bending along the y-axis is mainly seen. b) x-strain map at the same moment. X-strain is mainly compressive. c) y-strain map at the same moment. Y-strain is mainly expansive. The most prominent changes between the different mouthpiece designs (figs. 4 & 7) are seen in the x-strain map.

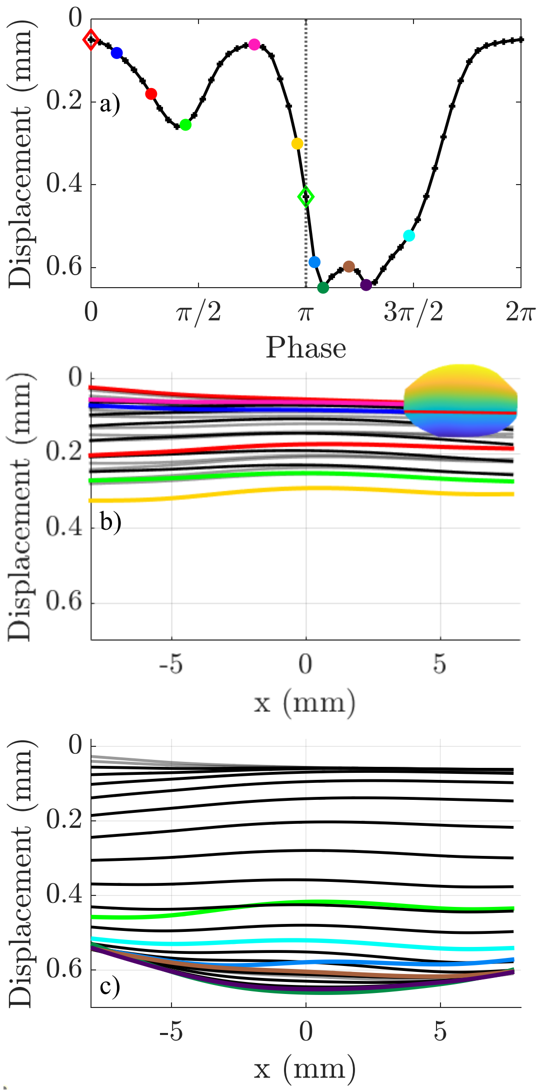


Figure 9: **Concept replica with low ridge** (a) displacement curve as a function of time of the middle point. The black dots indicate the 50 equidistant measurement points. The red and green diamond respectively indicate the start of the first and second half of the vibration phase. The colored dots indicate time points in the vibration cycle on which data will be presented in more detail. (b) displacement as a function of time for the first half of the vibration cycle for x cross-sections. Location of the cross-section is shown in the inset. (c) displacement as a function of time for the second half of the vibration cycle for x cross-sections. Due to the added ridge in the mouthpiece, less bending of the reed is observed compared to figure 5.

Figure 9 shows cross sections for the 50 time points along the x-direction taken at 3.5 mm from the tip. The displacement as a function of time for a point chosen on the visible part of the reed is also shown (figure 9a). Though the vibration cycle shows a similar as in figure 5, subtle differences can be seen when the reed is near the mouthpiece. When the mouthpiece with the low ridge is measured, a bounce effect of the reed is observed. When the reed is at its most downward position (dark green dot), it slightly opens again and then closes. The amplitude of the bounce is 50 micrometers. This bouncing effect was not observed in the original Concept and replica design. The cross-sections along the x-axis show less bending of the reed in the middle. The reed is now 0.1 mm lower at its center point compared to its outer edges. Although the ridge is lower than the reed, it does influence reed motion.

Figure 10 again shows the x-strain and y-strain for x-cross-sections at the selected time points (Fig. 9). Compared to the x-strain cross-sections measured on the original Concept mouthpiece (Fig. 6), lower x-strain values are measured at the reed middle for time points where the reed is near the mouthpiece. The x-strain value at x = 0 is twice as small on the Concept replica with the low ridge as for the original Concept. The maximum y-strain values measured on the Concept replica with the low ridge are similar to those measured on the original Concept mouthpiece. The outer sides of the cross-sections measured on the Concept replica with the low ridge show higher positive y-strain values. Starting from x = 0 mm and moving away from the middle, the values of y-strain decrease steeper to negative values on the original Concept mouthpiece.

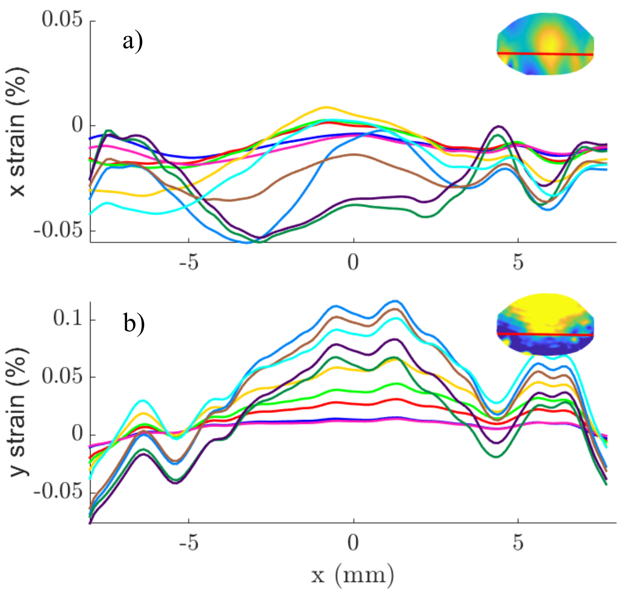


Figure 10: **Strain along a section parallel to the x-axis for the Concept replica mouthpiece with the low ridge** (a) x-strain at the time points indicated in figure 5. (b) y-strain for the same time points. Y-strain increases most at the center of the reed.

3.2 Subjective evaluation

The player gave the 3D printed Concept replica a score of 5 for playability and 4.5 for sound quality. She remarked that, in the very high notes, the sound was a bit sharper. The player found the high ridge replica to be totally unplayable and gave a score of 0. The player remarked that only with major effort could some sound be produced. It was not possible to play loud on the high ridge replica. For sound quality, the musician gave a score of 0 as well. The mouthpiece performed poorly over the entire registers, with many reed squeaks. The mouthpiece with the 0.2 mm lower ridge scored 3 for playability and 2 for sound. The lower notes till G/G# were playable reasonably well. From A3 on, the playability was considerably worse than on the original Concept. The overall sound was sharper, especially in the higher notes.

4. DISCUSSIONS & CONCLUSIONS

Stroboscopic DIC allows to measure full field deformation and bi-axial strain on a vibrating saxophone reed. When the reed touches the side rails of the mouthpiece, bending and strain in the x-direction occur, and the level of the center of the reed becomes lower than the side rails. When this bending is inhibited by placing a thin ridge under the reed, the mouthpiece becomes unplayable. When the top level of the ridge is just 0.2 mm below the level of the side rails, playability returns. These data show that bending along the direction perpendicular to the reed fibers is essential to saxophone reed dynamics and should be considered in reed vibration modeling. The data also show that reed elasticity parameters in the x-direction may be just as important as read elasticity along the length of the reed fibers to obtain correct reed vibration.

Acknowledgments

We gratefully acknowledge the support of J. Cottier (Henri Selmer, Paris, France) for placing the test mouthpieces at our disposal and M. Kortschot and S. King (Légère Reed, Barrie, Ontario, Canada) for placing reeds at our disposal. Further, we thank the professional saxophonist for her valuable input in the subjective evaluation of the different mouthpiece designs.

5. REFERENCES

[1] F. Pinard, B. Laine, and H. Vach, “Musical quality assessment of clarinet reeds using optical holography,” *The Journal of the Acoustical Society of America*, vol. 113, no. 3, pp. 1736–1742, Mar. 2003, doi: 10.1121/1.1543586.

[2] A. Ricardo da Silva, G. P. Scavone, and M. van Walstijn, “Numerical simulations of fluid-structure interactions in single-reed mouthpieces,” *The Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. 1798–1809, Sep. 2007, doi: 10.1121/1.2759166.

[3] F. Avanzini and M. van Walstijn, “Modelling the mechanical response of the Reed-mouthpiece-lip system of a clarinet. Part I. A one-dimensional distributed model,” *Acta Acustica united with Acustica*, vol. 90, no. 3, pp. 537–547, 2004.

[4] E. Ukshini and J. Dirckx, “Strain distribution on vibrating synthetic and natural saxophone reeds measured with digital image correlation,” presented at the Fourth Vienna Talk on Music Acoustics, University of Music and Performing Arts, Vienna, Austria, 2022, p. 035001. doi: 10.1121/2.0001640.

[5] A. Munoz, B. Gazengel, and J. P. Dalmont, “``In vivo'' and ``in vitro'' characterization of single cane reeds,” *SMAC*, 2013.

[6] A. Hirschberg, J. Gilbert, A. P. J. Wijnands, and A. M. C. Valkering, “Musical aero-acoustics of the

clarinet," *J. Phys. IV France*, vol. 04, no. C5, pp. C5-559-C5-568, May 1994, doi: 10.1051/jp4:19945120.

[7] Y. Shi, G. P. Scavone, and da Silva, Andrey R., "LBM SIMULATION OF THE QUASI-STATIC FLOW IN A CLARINET," *Proceedings of the Third Vienna Talk on Music Acoustics*, no. Third, 2015.

[8] T. Kobata and T. Idogawa, "Pressure in the mouthpiece, reed opening, and air-flow speed at the reed opening of a clarinet artificially blown.," *J. Acoust. Soc. Jpn. (E), J Acoust Soc Jpn E*, vol. 14, no. 6, pp. 417–428, 1993, doi: 10.1250/ast.14.417.

[9] V. Chatziioannou and M. van Walstijn, "Estimation of Clarinet Reed Parameters by Inverse Modelling," *Acta Acustica united with Acustica*, vol. 98, no. 4, pp. 629–639, Jul. 2012, doi: 10.3813/AAA.918543.

[10] E. Ukshini and J. J. J. Dirckx, "Influence of lip position, lip force and blowing pressure on the tuning and playability of an alto saxophone mouthpiece," *Applied Acoustics*, vol. 199, pp. 109011–109020, Oct. 2022, doi: 10.1016/j.apacoust.2022.109011.

[11] M. E. McIntyre, R. T. Schumacher, and J. Woodhouse, "On the oscillations of musical instruments," *The Journal of the Acoustical Society of America*, vol. 74, no. 5, pp. 1325–1345, Nov. 1983, doi: 10.1121/1.390157.

[12] A. H. Benade, *Fundamentals of musical acoustics*. New York: Dover Publications, 1990.

[13] E. Ukshini, P. Muyshondt, and J. J. J. Dirckx, "Blowing pressure stabilization method for the artificial excitation of reed instruments (L)," *The Journal of the Acoustical Society of America*, vol. 153, no. 2, pp. 921–924, Feb. 2023, doi: 10.1121/10.0017167.

[14] E. Ukshini and J. J. J. Dirckx, "Setup for musical reed instrument testing using constant phase stroboscopic imaging," *Opt. Eng.*, vol. 60, no. 04, p. 044105 (11pp.), Apr. 2021, doi: 10.1117/1.OE.60.4.044105.

[15] A. Joshi, S. Kale, S. Chandel, and D. Pal, "Likert Scale: Explored and Explained," *BJAST*, vol. 7, no. 4, pp. 396–403, Jan. 2015, doi: 10.9734/BJAST/2015/14975.

TROMBONE RANGE PROFILE AND VOICE RANGE PROFILE

Peter Pabon
Institute of Sonology, Royal Conservatoire, The Hague
p.pabon@koncon.nl

Wilco Kamminga
Master Student Trombone, Amsterdam Conservatorium, member Dutch Ballet Orchestra
wilcokamminga@hotmail.com

Geert Berghs
Associate Professor, Amsterdam Conservatorium
info@geertberghs.nl

ABSTRACT

Apart from the obvious differences, there are interesting parallels between the trombone and the voice. Both instruments are air driven by lung pressure and excited by a valve mechanism of vibrating lips, with further downstream a tube or tract that can be considered to act like a filter. In this study, we used spectral Voice Range Profile (VRP) recording [1],[2] to characterize the similarities and differences in acoustic features in the singing voice of a professional baritone and a trombone. Selected maps are presented; together they demonstrate the differences in sound characteristics, that we try to relate to differences in the sound production in both instruments. Maximum sound levels for the trombone were not dependent on the fundamental frequency. In the trombone, we found considerably lower jitter values, when compared to the voice. Several metrics characterize the trombone waveform as less complex, consisting mostly of a single pulse. The spectrum balance and H2/H1 parameters mark the sudden introduction of high frequency energy around 100 dB SPL over the whole tonal range of the trombone.

1. INTRODUCTION

An essential part of trombone lessons is based on parallels with the singing voice. “When you can sing it, you can play it” is often said by trombone teachers. That there is some benefit from singing lessons for a trombone player might be evident, but it is not so easy to quantify this in research. The voice and the trombone are rather different instruments. Their physical dimensions and structure, their mechanical control, and the output levels over their different pitch settings (range) are obviously different. However, both instruments are driven, or excited, by a mechanism of vibrating lips or vocal folds. The airflow plays an important role, but with rather different pressure levels. It is intriguing to compare the sounds of both instruments [3] [4][5] and to speculate which similarities in sound characteristics could be attributed to similarities in sound production mechanisms.

A Voice Range Profile (VRP) or voice map [1] [2], presents a systematic overview of how the acoustic quality of the voice transforms when the sound production conditions change. It aims to get an all-inclusive recording that collects the complete gamut of spectral and time-based qualities that an instrument is capable of. It thereby gives information on the underlying mechanisms. Applying this mapping paradigm to the trombone and creating a Trombone Range Profile (TRP), is the aim of this study, where the VRP of a professional baritone singer is taken as a reference.

An important reason to apply the VRP recording technique is its ability to assess small changes in the functioning of the voice. This discriminative ability is determined by two independent parameters that are also the coordinates of the map, the f_o and the SPL. These are highly defining for the overall timbre or quality of the sound production mechanism. The f_o/SPL space defines its own distinctive continuum, revealing connections between voice samples that may be acquired at very different instances in time. Testing how the sound quality changes with SPL implies an evaluation of the instrument’s ability to build more high frequency energy when the sound pressure increases. Another aim in this study is to find the specific sound levels at which high frequency energy is introduced in the spectrum, and if we can recognize specific spectral transitions over the tonal range.

2. METHOD

Voice Profiler (VP 5.2) spectral, developed by the author PP, was used to record the TRP and VRP. This interactive recording system uses a special dual microphone headset [6]. The VP recording software, continuously analyses and compares the sound signals coming from both a far microphone and a microphone close to the mouth (see Figure 1). The far microphone of the VP headset maintains the reference for the SPL calibration, and is positioned at 30 cm distance from the mouth, which is the standard distance for VRP recording [4]. The SPL calibration for the signal picked up by the close microphone is continuously adjusted to the levels measured by the far microphone at the 30-cm standard distance. Additionally, there is a continuous monitoring of the time delay between both microphone signals and automatic checks on saturation, clipping and hum. With these extra checks the system is able to automatically select the sound source for the recording.

To record the TRP, this dual microphone headset was mounted on the bell (Figure 1). The close microphone was positioned ca. 7 cm from the exit plane determined by the rim of the cone. The far microphone of the VP headset, was positioned ca. 37 to 40 cm further from the rim. It is estimated that this 7 to 10 cm extra in ‘mouth’-microphone distance lowered the TRP levels about 1.5 dB compared to the levels based on the 30-cm standard used for the VRP.

Figure 1 Headset microphone placement with the singer and with the trombone.

The recording system provides data on the extremes of the vocal/trombone range, designated by the ‘contour’ of the area, and a set of parameters, characterizing temporal and spectral features of the output signal. These parameters are designed to characterize the sound production mechanism typical to the human voice. To what extend these parameters are applicable to characterize the trombone is still unknown. For the following parameters or voice metrics, VRP maps will be presented and discussed: *jitter* & *crest factor*, (derived from the time waveform) and *spectral cresting*, *spectrum balance*, *H2/H1 level balance*, (derived from the average narrowband FFT spectrum stored per cell).

The *jitter*, or period-to-period fluctuation in duration, expresses the regularity or periodicity of the fundamental in the output signal. It measures the standard deviation in the duration of successive periods using the negative pressure peak as a period marker. Due to the adaptive interpolation scheme implemented, the system is able to detect period duration differences in the order of a microsecond.

The *crest factor* tells how much one single cresting peak in the waveform determines the overall (RMS) energy in one period.

Spectral cresting describes how much the strongest harmonic in the spectrum is raising above the rest of the spectrum, and indicates if specific resonances of the vocal tract or trombone may selectively boost other harmonics than the fundamental (the first harmonic).

Spectrum Balance (SB) is a gross indicator of the steepness of the spectrum contour’s slope. For the voice, a flatter slope indicates a more powerful or abrupt excitation of the vocal tract and a sharper sound.

H2/H1 level balance is an indicator of the length of the closed phase that typically occurs in the glottal flow signal for the voice. With a pressed voice the level of the fundamental (H1) is generally lower.

The singer, of whose voice the VRP was made, is a professional baritone, graduated with a Master Degree and active as a soloist as well as in professional ensembles. The trombone player (WK) is a Master student at the Amsterdam Conservatorium and orchestra player (author WK).

The instruction given to the singer in making the VRP was to sing the vowel /a/ on every tone in his vocal range, starting at a comfortable volume and from there sing as softly as possible as well as at his loudest. The trombone player followed the same procedure: he started on every tone in his range at a comfortable volume, from where he made decrescendos and crescendos into the extremes, while keeping the suggestion of the vowel /a/ in the mouth. This way the full SPL range was covered semitone-by-semitone, acquiring a minimal density (time integration) of at least 0.10 seconds of sound material for each cell, of 1 decibel x 1 semitone width. For the voice, the samples were made avoiding falsetto register, aiming at an equalised voice quality.

The VRP of the same player-trombone combination was re-recorded three times using the same VP recording software, but in three different locations using three different microphone amplifier versions. This allows an additional check for variability with repeated measurements.

3. RESULTS

3.1 TRV and VRP contour, Phonation Density

Figure 2 shows the Trombone Range Profile (TRP) and the Voice Range Profile (VRP), that sketch the contour of the range. The colour represents the phonation density, the accumulated time spent at each storage cell in the map.

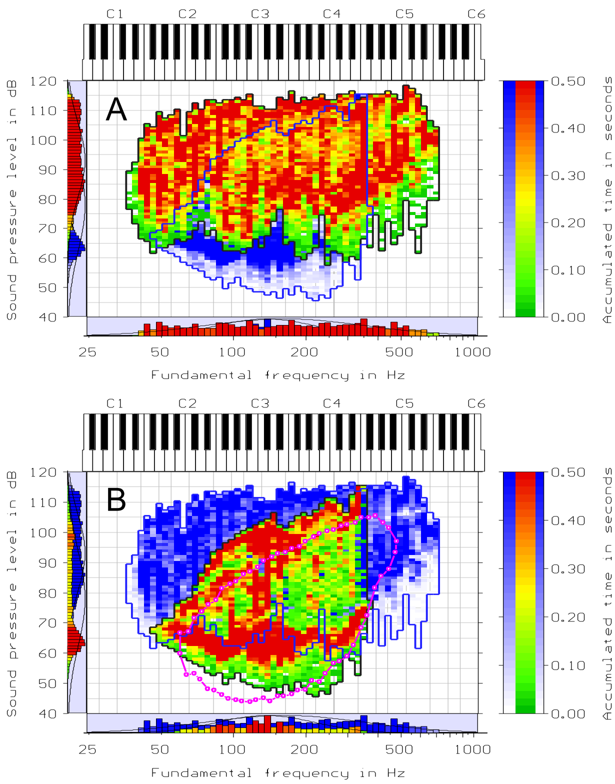


Figure 2. Phonation density metric (accumulated time per cell). Panel A: TRP area covered by the trombone. Panel B: VRP area covered by a professional baritone singer using modal voice (M1) only. The contour (solid blue line) and density (blue

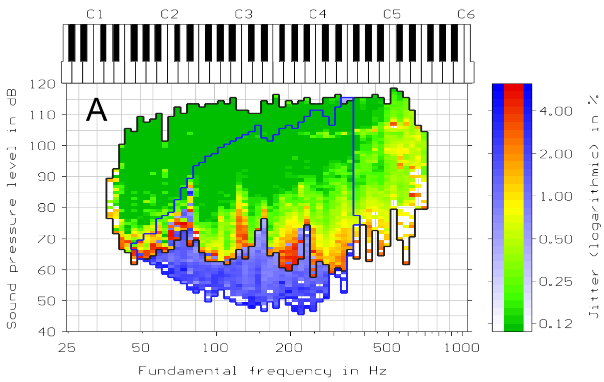
shading) for the complementary instrument reappear at the background. The pink curve (panel B only) shows the average contour [9] for 7 trained voices all using M1 only.

This density metric is standardly visualized during the interactive recording and it is the logical metric to present first here, as it delineates the base area for the feature maps presented later. The trombone is capable to produce tones over a wide tonal range of more than four octaves (usable E1 to F5, 41..700 Hz), with extreme sound levels topping 115 dB SPL. The modal voice of the chosen baritone covers about 3 octaves. As typical for the voice, the maximum sound levels reached, increase with the fundamental frequency [8]. Compared to the average contour for (N=7) professional male voices recorded with the same system (pink line), this professional baritone reaches ca. 10 dB higher maximum sound levels all over his tonal range. Still, these levels don’t compare to the levels reached by the trombone for the same pitch. Only on the top of his vocal range in M1 (tone E4), the baritone was able at high effort to touch the same extreme SPL as the trombone. The red zones on the colour map for the phonation density indicate that the singer spent considerable time searching either for the lowest levels or for the highest levels at a given pitch. While doing meticulous crescendos and diminuendos, the trombone player aimed to spend approximately the same amount of time per cell, leading to a more even coverage of the vertical range.

The total pitch range of the trombone was covered by applying different sliding regimes with at some points adaptations in embouchure and tube length or valve settings between two successive tones in the range, especially when the tube could no longer be lengthened further to lower the pitch. The largest adaptation is found around B1, where also a dip in the maximum contour is seen. The contour of the TRP does not show a gradual increase in maximum levels with increasing fundamental frequency, and does not show the typical oval shape characteristic to the VRP.

3.2 Jitter

Figure 3 shows the period-to-period jitter for the trombone and baritone.



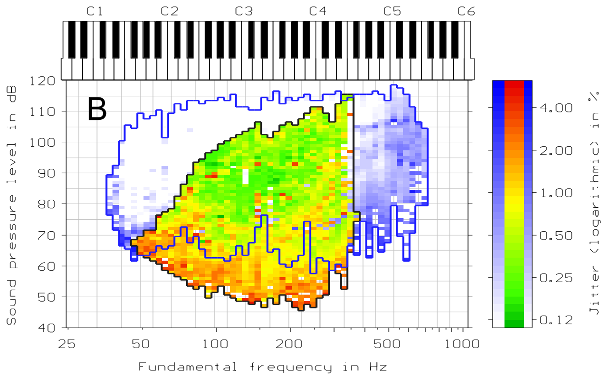


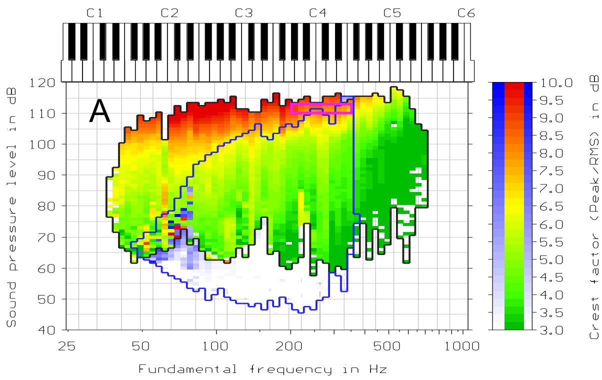
Figure 3. Period-to-period jitter (log) in percent for the trombone (panel A) and the baritone voice (panel B). For the paired instrument the contour is reproduced (solid blue line) and its jitter is coded with a blue shading at the background.

Over a large part of the trombone range, the jitter remains below 0.125 % of the period duration (Figure 3A, dark green colour). Such extreme low jitter values are infrequently observed with the paired professional baritone voice that was selected for this comparison, but also rarely with the voice in general. The jitter values for this baritone voice fluctuate around 0.5% (Figure 3B, light green colour).

3.3 Crest factor

The map for the trombone in Figure 4A shows a red coloured zone above 100 dB SPL, that indicates the crest factor (Peak/RMS level ratio) of the audio signal comes close to 10 dB. This means that at these extreme sound levels the waveform has become extremely spiky.

With the voice operating at high SPL (see Figure 4B), the crest factor also increases, but it generally settles at a lower ratio at around 7 dB. The red spot located around C2 and 75 dB SPL marks phonations where the singer used a pressed voice with a throaty quality while searching for the lowest tones in his range. With this voice quality, the audio waveform becomes more spikey, which the high crest factor values indicate.



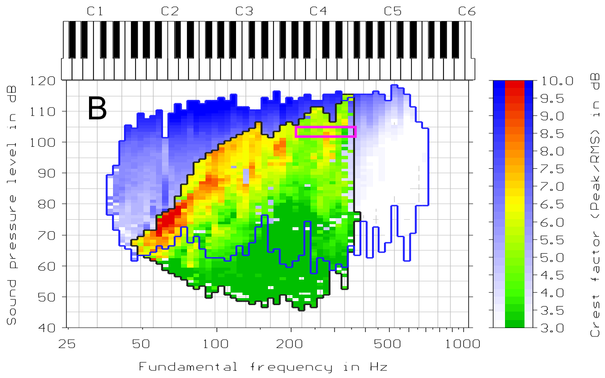
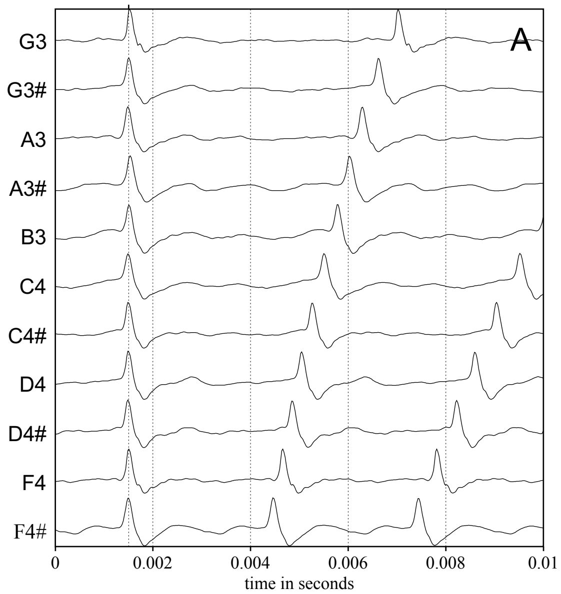


Figure 4. Audio crest factor (Peak/RMS level ratio) in dB for the trombone (panel A) and the baritone voice (panel B). The pink rectangles already indicate selections of the range that will be inspected in more detail in the next paragraph.

To clarify the origin of the differences in crest factor between the trombone and the singing voice, Figure 5 shows a series of short samples of the periodic waveform, for both instruments, observed at successive tones in the range.

Each fundamental period of the trombone (see Figure 5A) starts with a single and sharp peak of positive polarity. This peak is followed by a less extreme and gradually widening negative dip that smoothly extends into a ripple of a much lower amplitude. This pattern with the transient and the ripple that follows remains constant, independent of the pitch.



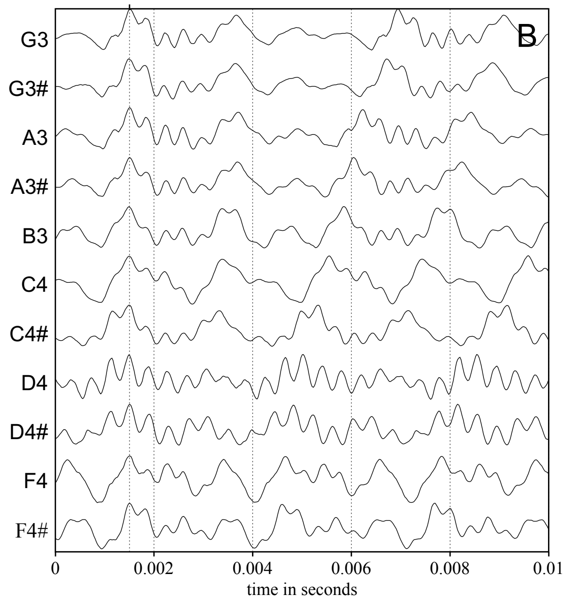


Figure 5. Audio waveform samples for 11 successive semitones (G3 to F4# / 182 to 337 Hz) appearing at 110 dB SPL for the trombone (A), and at 103 dB SPL for the baritone voice (B). Waveforms were synced by hand at 0.0015 sec at the top of their highest periodic wave crest.

For the voice samples (see Figure 5B), it is less clear what could be considered a starting point of a period, when compared to the trombone. In the voice, we see no recurrent waveform pattern to be associated with an excitation instant, typically the moment of vocal closure. Note that with every next semitone, the response to the excitation of the vocal tract may result in a slightly different oscillatory pattern. The general pattern shows the combined result of mainly F1, F2, and the singers formant, resonating together. However, at different fundamental frequencies, the formant that gains the upper hand may change. For instance, with tones D4 and D4#, the faster ripple (3 cycles in a millisecond) is dominating, which indicates that a singers formant around 3 kHz (3000 cycles in a second) largely controls the waveform. Generally, this tail of combined formant resonances smears out the energy of the initial transient over the fundamental period. This leads to a lower crest factor, as the average energy level (RMS) over the waveform will not defer largely from the peak level.

3.4 Spectral cresting

If specific resonances of the vocal tract or trombone may selectively boost other harmonics than the fundamental (the first harmonic), then the question is how much the strongest harmonic in the spectrum is raising above the rest of the spectrum. This aspect is represented in Figure 6.

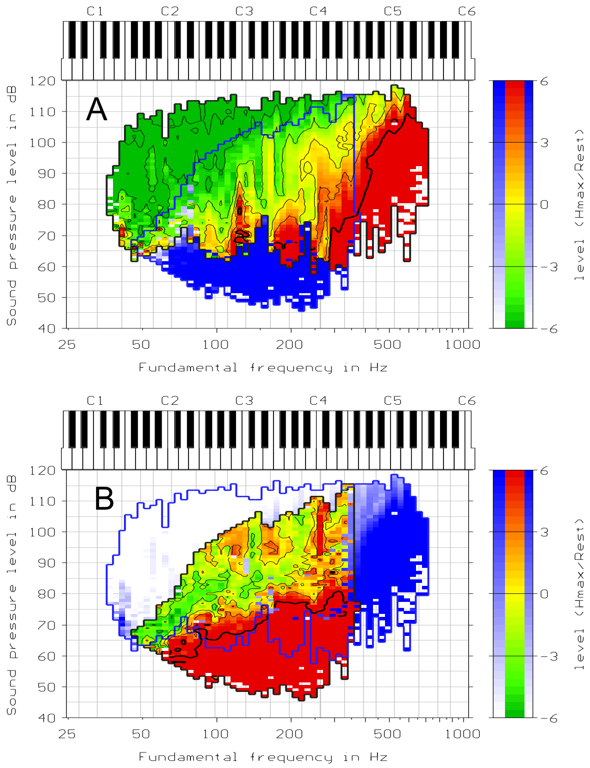


Figure 6. level ratio of the strongest harmonic in the spectrum and the overall energy in the rest of the spectrum in dB. Trombone (panel A) and the baritone voice (panel B).

Both maps show a red area along the lowest levels in the range. This red zone widens with increasing f_0 and marks the basic spectral configuration where the fundamental is still the leading spectrum component. For the voice and the trombone, operating at higher sound levels, generally a higher harmonic will be holding the highest power in the spectrum. Which harmonic this is, is not represented in the graphs. The green coloured zone in Figure 6A does tell us though, that whatever the harmonic number, it will never stick out far above the rest of the spectrum. The strongest harmonic will always have close competition nearby e.g. it will never designate a distinct resonance. A smooth spectrum envelope is characteristic for a periodic waveform consisting of a single pulse. Thus, this spectral feature reconfirms the earlier observations for the trombone waveform as being predominantly a single pulse.

With the singing voice (Figure 5B) operating at levels above 90 dB SPL, there are about three red blobs: one around A2/B2, a second around E3/F3, and a third marked by a vertical line at C4. These three red zones mark where a single harmonic, other than the fundamental, is sharply peaking in the spectrum, likely by interacting with a formant resonance that matches in frequency. Although interesting, the question which harmonic is interacting with which formant is not further explored in this context.

3.5 Spectrum Balance

In both instruments, we see a tendency of more energy spreading into the higher harmonic frequencies, with higher sound levels. Figure 7 reports the level balance for the band below 1.5 kHz in relation to the band above 2 kHz for both instruments. A red colour here represents a relatively high energy above 2 kHz in the output spectrum.

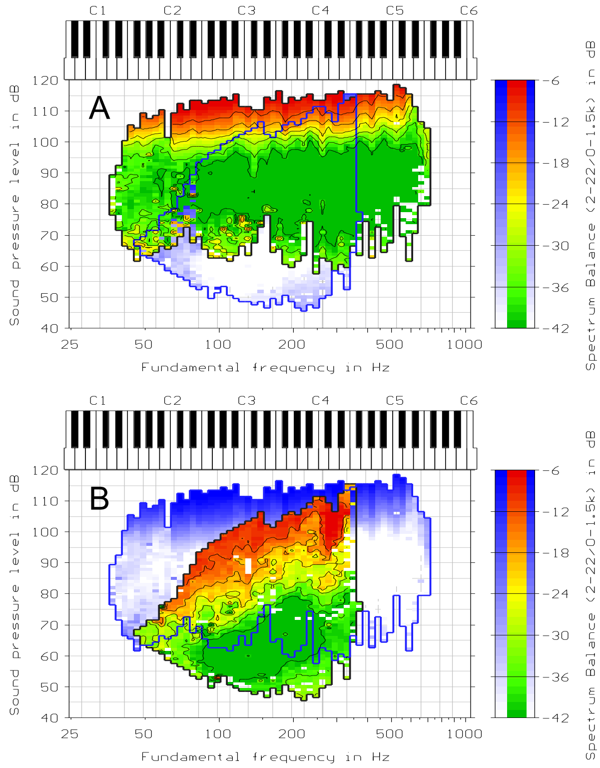
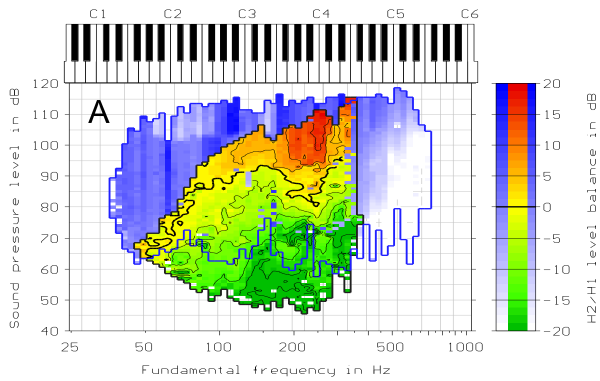


Figure 7. Spectrum balance for the trombone (panel A) and the baritone voice (panel B).

The trombone is able to produce sounds at much higher sound levels. The ability to drive this energy up, into the higher harmonics above 2 kHz in the spectrum, appears above 100 dB SPL (the red colour in Figure 7A). For the voice (map B), the balance is rapidly tilting to the high frequencies, at sound levels already around 70 to 80 dB SPL, depending on the connected fundamental frequency.

3.6 H2/H1 level balance

The H2/H1 level balance for the voice is assumed to be related to the change in length of the closed phase in the glottal flow signal [2]. The distribution seen for the voice, now appearing first as the top panel A in Figure 8, is typical for this specific voice: in repeated recordings of this voice, more or less the same distribution pattern will reappear. The distributions for the H2/H1 metric vary widely, seen from singer-to-singer, even for voices in the same voice group [8].



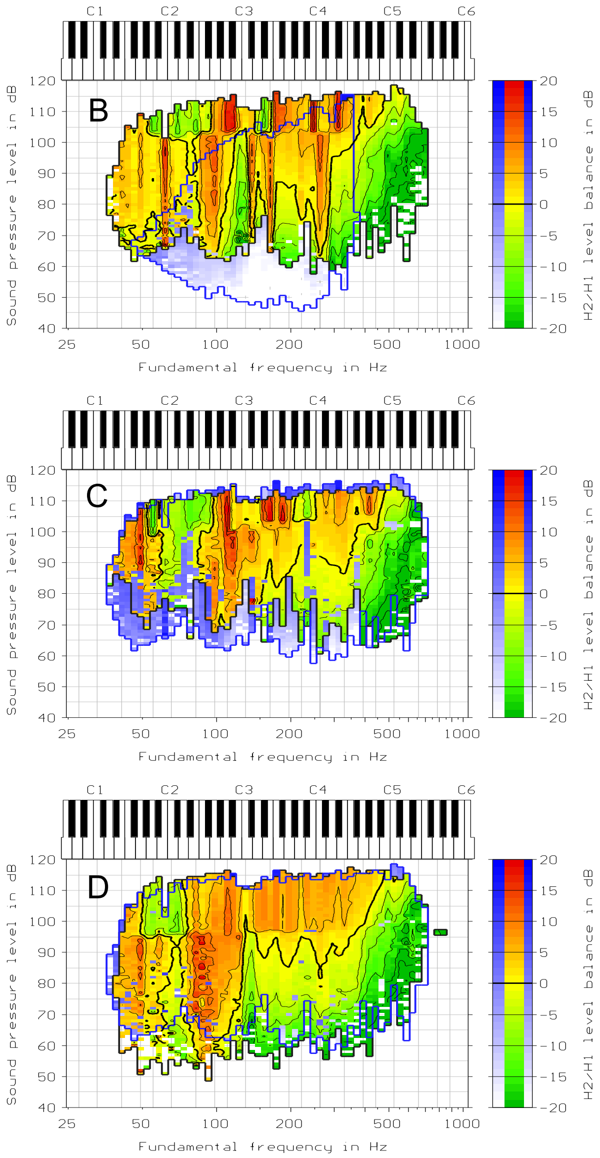


Figure 8. H2/H1 level balance for the selected baritone voice (panel A) and the same trombone recorded 3 different instances at different locations and with different hardware versions (panels B, C & D).

The three recordings made with our trombone player, portray largely identical spectral and waveform features, and are therefore not shown, except for this metric. All three different recordings are shown in Figure 8B, C and D. They all show that there is a H2/H1 balance inversion happening over a shared horizontal line somewhere around 100 dB SPL, however in different directions. For instance, at E3, in Figure 8B the balance flips from orange (+5 dB) to green (-10 dB) with increasing SPL. But in Figure 8C we see a flip at E3 from orange (+5 dB) to red (+10 dB). In Figure 8D we see the same phenomenon at E3 from yellow (0 dB) to orange (+5 dB). The direction of the H2/H1 balance change with SPL may be opposite for the same pitch, at different recording occasions, and seems to resettle with each onset of a new tone. Although strange and intriguing, this result is not just an error. It may have an explanation that could be very relevant here.

4. DISCUSSION

The VRP contour is typically an oval shape, that reflects the strong covariation that generally is seen between f_0 and SPL. There are no vertical oriented sudden transitions or discontinuities in quality, although the area was recorded semitone-by-semitone.

The TRP contour is basically a rounded rectangle, and the tonal range exceeds 4 octaves. Much higher sound levels were recorded over a wider part of the range, when compared with the VRP. A gradual increase in maximum SPL with increasing f_0 was not observed. This suggests that all over the tonal range, the same driving pressure generates tones with comparable sound levels.

At B1 a dip in the maximum contour was seen that could be explained by a large adaptation in the tube length.

Over the full range of possible sounds that can be produced by the trombone, the measured period-to-period jitter reaches extremely low values, typically levelling at values below 0.125%. For the compared baritone voice, or with the voice in general, the lowest jitter values found still vary around 0.5% or more. This strong contrast demonstrates how with the trombone, the filtering of the connected tubing totally rules the fundamental periodicity of the lip vibration in the mouthpiece. Contrary, the filtering by the vocal tract does not have this kind of influence. It is reasonable to assume that both the lips of the mouth and the vocal folds are equally able to perform a complex vibration pattern with a certain freedom to deviate in their periodic motion. The low jitter suggests that with the trombone, the lips vibrating in the mouthpiece are more restrained in this freedom.

Comparison of the crest factor for the voice and the trombone reveals a fundamental difference in complexity of the waveforms. For the trombone, each period contains a single cresting peak. This peak sharpens-up with increasing sound levels. For the voice, the formant resonances together smear out the energy of the initial transient over the fundamental period.

Testing for spectral cresting revealed a predominantly flat trombone output spectrum without sudden fluctuations in harmonic strength, due to interactions with formant-like resonance modes. Nowhere in the range of sounds producible by the trombone, a frequency selective filtering mechanism was favouring one harmonic only, apart from the fundamental. In the singing voice, as expected, harmonics selectively peak or crest when interacting with different formants.

Both instruments demonstrate that with higher sound levels, more energy will be spreading into the higher harmonic frequencies. With the voice the exciting pressure transient is the result of a sudden breaking of the flow by the closure of the vocal folds. With the trombone, extra energy appearing above 2 kHz is the result of a different mechanism: the transient sharpens up due to speed differences, depending on the pressure distribution over the running wave, eventually leading to a shock wave [10][11][12][13]. This process relies on the smoothness of the filter, seen over the frequency range, but also spatially seen over the geometry of the tubing. This is a more powerful mechanism in terms of the SPL reached. Maximum power levels reached with the voice are considerably lower, but still relatively high. For the voice, high frequency energy emerges at a lower SPL, and as a bonus the vocal tract filter has the freedom to be non-uniform and to dynamically vary its spatial geometry.

The observations on the H2/H1 balance suggest that the excitation mechanism controlling this parameter is essentially different for the voice and the trombone. For the voice, the interpretation is linked to the introduction of a closed phase, that may vary in duration. The tilting of the H2/H1 balance in the trombone, which occurred in opposite directions with more driving pressure, we cannot explain.

The SPL, where the flipping of the H2/H1 balance is seen, remains more or less constant with f_o, leading to an imaginary horizontal line around 100 dB SPL in the TRP. The same horizontal line is seen in the Spectrum Balance, where it marks the sound level from where energy above 2 kHz is introduced. As this line follows the maximum SPL contour over the total tonal range, this again confirms the earlier observation that tones with comparable sound levels and quality relate to the same driving pressure in the trombone.

5. CONCLUSIONS

Comparing the sound of the singing voice with the trombone, using the VRP/TRP mapping technique, revealed substantial differences in sound characteristics, expressed in the contour, the jitter, the crest factor, and several spectral parameters. Our first application of this technique brought intriguing graphical representations, that present a new perspective on the trombone sound, and sound production mechanism. Such as that a trombone player seems to apply the same driving pressure over his whole tonal range to realise tones of different pitch but with comparable sound level and spectral qualities.

6. REFERENCES

[1] S. Ternström, P. Pabon, "Voice Maps as a Tool for Understanding and Dealing with Variability in the Voice," in Porto, 2009, pp. 213-218.

[2] P. Pabon, S. Ternström, "Feature Maps of the Acoustic Spectrum of the Voice," in J. Voice 2020, 34,161.e1–161.e26.

[3] A.H. Benade, Fundamentals of Musical Acoustics. Courier Corporation, 1990.

[4] X. Pelorson, R. Msallam, J. Gilbert and A. Hirschberg, "Fluid dynamic aspects of human voice and brass instruments: implications for sound synthesis," in Proceedings ICA/ASA, Seattle. 1998.

[5] T. Hézard, V. Fréour, R. Caussé, T. Hélie, and G. P. Scavone, "Synchronous multimodal measurements on lips and glottis: comparison between two human-valve oscillating systems," in Acta Acustica united with Acustica, 100(6), 2014, pp 1172-1185.

[6] T. Printz, J.R. Sorensen, C. Godballe, Å.M. Grøntved, "Test-Retest Reliability of the Dual-Microphone Voice Range Profile," in J. Voice 2018, 32, pp. 32–37.

[7] H.K. Schutte, W. Seidner, "Recommendation by the Union of European Phoniatricians (UEP): Standardizing voice area measurement/ phonetography," in Folia Phoniatr. Logop. 35(6), 1983, pp 286-288.

[8] P. Pabon, "Mapping Individual Voice Quality over the Voice Range: The Measurement Paradigm of the Voice Range Profile." Ph.D. Thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2018.

[9] P. Pabon, S. Ternström, and A. Lamarche, "Fourier descriptor analysis and unification of voice range profile contours: method and applications," in Journal of Speech Language and Hearing Research, 54, 2011, pp. 755–776.

[10] A. Hirschberg, J. Gilbert, R. Msallam, and A.P.J. Wijnands, "Shock waves in trombones," The Journal of the Acoustical Society of America, 99(3), 1996, pp. 1754-1758.

[11] M. Campbell, J. Gilbert, and A. Myers, The science of brass instruments. Switzerland: Springer 2021.

[12] D.M. Campbell, J. Gilbert, and A. Myers, (December). "Brassiness and the understanding of brass instruments," in Forum Acusticum 2020, pp. 3169-3175.

[13] A. Myers, R.W. Pyle Jr, J. Gilbert, D.M. Campbell, J.P. Chick and S. Logie, "Effects of nonlinear sound propagation on the characteristic timbres of brass instruments," The Journal of the Acoustical Society of America, 131(1), 2012, pp. 678-688.

THE REFLECTION WAVE AND LOUDNESS VARIATIONS IN HARPSICHORDS

Geert Karman
geert@piparte.com

ABSTRACT

The motion of a plucked harpsichord string during the time it is lifted at the plucking point is analysed with a numerical model in order to track the wave associated with the attack of the plectrum on the string. Depending on the timing of the arrival of the reflections from the fixed ends back at the plucking point, the force that the string exerts onto the plectrum can increase or diminish, leading to variations in the maximum amplitude at the moment of release. This timing is determined mainly by the frequency of the note. Based on simulations with the numerical model, a diagram is compiled that summarises the practical possibilities for obtaining loudness variations over the keyboard range of the harpsichord. Understanding the workings of the reflected attack wave also allows for the compilation of a list of characteristics of the construction and regulation of a harpsichord that can conceivably influence a player's ability to produce loudness effects on a particular instrument.

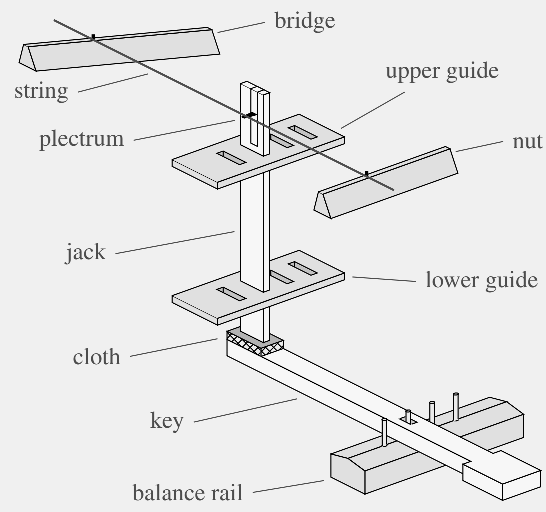


Figure 1. *Schematic of the harpsichord mechanism.*

1. INTRODUCTION

In a harpsichord the plucking of a string is achieved through a mechanism that holds a plectrum. (See Figure 1 for a schematic with the essential components.) The fixed distance between the string and this mechanism gives the harpsichordist restricted possibilities to achieve variations in loudness. During normal playing the volume of individual notes is constant, but by striking a key much faster a small increase in volume can be obtained. This effect has been attributed to the reflection of a wave that is generated during the contact between the plectrum and the string. Sometimes these variations in loudness are referred to as dynamics, but harpsichord players use this term more in connection with performance, where a variety of methods are used to make accentuations (such as articulation, the way of executing arpeggios and other means) rather than simply changes in the volume of individual strings. So to avoid any possible confusion, I will use the term "loudness variations". The wave that is generated during the contact between the plectrum and the string will be called the "attack wave".

Hall [1] was the first to mention reflection waves as a possible cause of variations in loudness. Griffel [2] studied the dynamics of plucking with a simplified mathematical model and found that there would be relatively large variations in loudness depending on the attack speed. In particular, he noted that when increasing higher attack speeds further there would first be an increase in loudness and then a drop-off to almost no sound. Giordano and Winans [3] tried to verify Griffel's results with an experimental setup consisting of a single key and string. They also developed an improved mathematical model for plectrum-string interaction. They found much less variation in loudness than Griffel, both experimentally and with their mathematical model. Also, although their mathematical model agreed with Griffel that for very high attack speeds the volume would drop to almost zero, their experiments did not confirm this. They too mention the reflection of the attack wave as being the cause for the small variations they found but did not investigate this. Perng [4] further improved on the mathematical model of Giordano and Winans and qualitatively confirms their results, but his model gives a better match with their experiments. He also investigates the effect of the position of the plucking point on the attainable maximum amplitudes with his numerical model and finds that the closer the plucking point is to the middle of the string, the higher the maximum amplitude.

Loudness variations have also been the subject of various experimental studies with real instruments, but the role of the reflection wave seems not to have been taken into consideration in these studies. Penttinen [5] used a harpsichord with Italian and Southern German characteristics to measure volumes and envelopes of harmonics for the C keys. He found small but audible variations in both volume and timbre for different striking velocities of the key. Another finding of his study was that the action of the harpsichord could enhance the impression of a stronger sound by exciting the case of the instrument, the "knock". MacRitchie and Nuti [6] conducted a similar study on a historical harpsichord (which had undergone a far-reaching restoration of the action) and found similar results. In addition to three registers with quills, the harpsichord in question also possesses a so-called "peau de buffle" register where the plectrums are made out of leather. This register was slightly better at producing changes in loudness. Paté et al. [7] did experiments where a robotic finger was used to press one key (G_4) with various plectrums. The instrument was made in 1989 to specifications of Flemish harpsichords of the 17th century but with only one keyboard and one register. They found no variation in string amplitudes with varying speeds of the key dip.

In this study I will use a numerical model to determine the role of the reflection wave in producing variations in the loudness of an individual, plucked string. I will next look into the implications this has for the possibilities to produce these loudness effects on a harpsichord. The general conditions under which variations in loudness can be achieved over the range of the keyboard will be summarised in what I call a loudness map. I will also compile a list of factors that might be expected to affect the possibilities for achieving a greater loudness in an individual instrument.

2. NUMERICAL MODEL

Consider the one-dimensional equation for transverse waves in a completely flexible string

$$\left(\frac{\partial^2 u}{\partial t^2}\right) = \frac{T}{\mu}\left(\frac{\partial^2 u}{\partial x^2}\right) + S_p \tag{1}$$

where x is the space coordinate along the string and t the time coordinate, T is the tension in the string and μ the linear density. S_p is a source term that represents the plucking action of the plectrum. This equation is to be solved for the transverse string displacement over time $u = u(x,t)$ in the domain [0,L], L being the string length.

Both ends of the string are fixed so the boundary conditions are

$$u(0,t) = 0, \;\; u(L,t) = 0 \tag{2}$$

At $t = 0$ the string is without any displacement and at rest, giving the initial conditions

$$u(x,0) = 0, \;\; \frac{\partial u}{\partial t}(x,0) = 0 \tag{3}$$

Instead of aiming for an accurate modelling of the plectrum, I will instead go for a simple approach and remove the plectrum as much as possible from the model. It is anticipated that the creation of an attack wave without accounting for the dynamics of the bending plectrum is enough to study the phenomena at issue. From the experiments in [3] and also from calculations with a quasi steady-state model in two dimensions [8] it was established that the force of the plectrum on the string can be approximately described as a linear increase over time

$$\begin{aligned} S_p &= \frac{F_{max}}{\mu}\frac{t}{t_r - t_0} \quad \text{for} \quad x = x_p,\;\; t_0 < t \le t_r \\ S_p &= 0 \qquad\qquad\quad \text{otherwise} \end{aligned} \tag{4}$$

where x_p is the location of the plucking point, t_r the moment where the string is released from the plectrum and t_0 the moment of first contact between the plectrum and the string. The attack time span (the time span during which there is contact between the plectrum and the string) is then $\Delta t_a = t_r - t_0$. F_{max} is the maximum force that the plectrum can exert on the string. This force is given by the strength of the voicing and is defined by the dimensions and shape of the plectrum. It is assumed here that the plectrum can support any force less than this maximum force but immediately releases the string when this force has been reached. At that moment the source term S_p is set to 0 again.

The proposed measure for the loudness is the energy that the string has at the moment of release, in line with the findings in [3]. The actual volume that is produced will evidently also depend on the characteristics of a particular instrument (string-soundboard coupling and radiation phenomena), but in this context we are interested in loudness variations for a single string and can thus use energy variations as a measure for loudness variations. The energy can be found from the amount of work W_s put into the string

$$W_s = \int_0^{u_{p,r}} F_p \mathrm{d}u_p \tag{5}$$

where F_p is the force from the plectrum on the string and u_p is the displacement of the string at the plucking point. The upper limit $u_{p,r}$ of the integral is the displacement of the plucking point at the moment of release. The energy will be stored partly as potential energy in the deviation of the plucking point and partly as kinetic energy contained in the movement of the string.

The implementation of the model as a computer program uses a finite difference method with a fourth order central difference scheme for the space derivative. The solution is advanced in time with a first order explicit Euler scheme. The boundaries were implemented using extra cells outside the domain with a fixed solution $u = 0$ to be able to calculate the space derivative with the central difference scheme.

3. SIMULATION CASES

In this section, four simulations will be presented that were done with the model. To keep these simulations realistic I have taken the lower manual of a French double manual harpsichord (Hemsch, 1751) as a reference. Three cases simulated plucking the strings for C_2 (two octaves below middle C), C_4 (middle C) and C_6 (two octaves above middle C). A fourth case had the same parameters as for C_4, but this time plucked in the middle of the string. This does not of course reflect a realistic situation for the Hemsch, but it was added to examine the influence of the position of the plucking point. Table 1 gives an overview of the string parameters that were used in each of the simulation cases.

Case	L (mm)	x_p (mm)	μ (g/m)	T (N)	freq. (Hz)	cycle (ms)
C_2	1695	202	1.392	61	62	16.1
C_4	700	150	0.444	54	249	4.02
C_4^*	700	350	0.444	54	249	4.02
C_6	184	92	0.322	43	993	1.01

Table 1. *String parameters for the four simulation cases.*

Table 2 gives the maximum displacements $u_{p,r}$ that were chosen as the basis of the plucking actions in the simulations. The values for the cases C_2, C_4 and C_6 were estimated with a caliper in a copy of the Hemsch instrument with what is considered a moderate voicing for a quasi steady-state release (where $\Delta t_a \to \infty$). The exact magnitudes of these values is not really important; they merely serve as a reference for the discussion in Section 7. The value for case C_4^* was chosen so that the work W_s performed on the string is equal to the amount of work for C_4 (except for some round-off error). The associated maximum plucking forces (for a steady-state release) can readily be found from an equilibrium of forces just before the release of the string

$$F_{max} = u_{p,r}T(\frac{1}{x_p} + \frac{1}{L - x_p}) \tag{6}$$

The corresponding amount of work W_s put into the string is then easily calculated by evaluating the integral of Equation (5), using the expression for the force that the plectrum exerts on the string for any given value of the deviation u_p of the string at the plucking point

$$F_p = u_pT(\frac{1}{x_p} + \frac{1}{L - x_p}) \tag{7}$$

Note	$u_{p,r}$ (mm)	F_{max} (N)	W_s (J $\cdot 10^{-3}$)
C_2	2.0	0.69	0.688
C_4	1.0	0.45	0.221
C_4^*	1.4	0.37	0.219
C_6	0.4	0.40	0.086

Table 2. *Maximum plucking point displacements $u_{p,r}$ with the corresponding plectrum forces and work W_s put into the string. These values are for quasi steady-state plucks ($\Delta t_a \to \infty$).*

For each of the cases presented above a series of calculations were performed for a range of attack time spans Δt_a. In all simulations the number of (equal sized) spatial cells was 1000. The time step sizes were chosen such that the Courant number was small enough for the simulation to be stable. Several tests with smaller time step sizes and/or spatial cells did not reveal any significant changes in the results.

4. LOUDNESS VARIATIONS

Figures 2, 3 and 4 show the amounts of energy that result from different attack time spans Δt_a for cases C_2, C_4 and C_6. In order to compare these results to the findings of Griffel [2], Giordano and Winans [3] and Perng [4] we have to bear in mind that in their cases they have the upward jack velocity as an independent parameter, whereas here the attack time span is the independent parameter. These two parameters are linked by an inverse relationship that depends on the model of the bending plectrum, so the comparison is not straightforward in all its details. What is clear however, is that there is qualitative agreement: for very small Δt_a (very high jack velocities) the amount of energy in the string becomes very small. When increasing Δt_a (diminishing the jack velocity) we gradually get to a maximum energy. Increasing Δt_a further then causes a first minimum of energy and subsequently a succession of maxima and minima, with decreasing amplitude as Δt_a grows, converging to the steady-state energy.

The use of the attack time span Δt_a instead of the jack velocity as an independent variable reveals that successive maximum energies occur at intervals for Δt_a that are the cycle time of a vibration apart. (Compare the distances between the vertical lines in Figures 2, 3 and 4 with the last column in Table 1.) This is already a strong indication that the maximum and minimum energies are linked to a wave in the string.

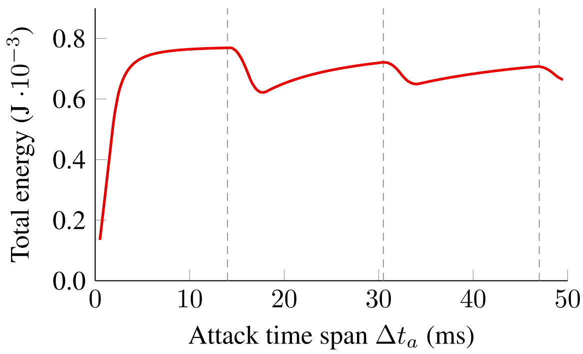


Figure 2. *Energy stored in the string for C_2 at release as a function of Δt_a.*

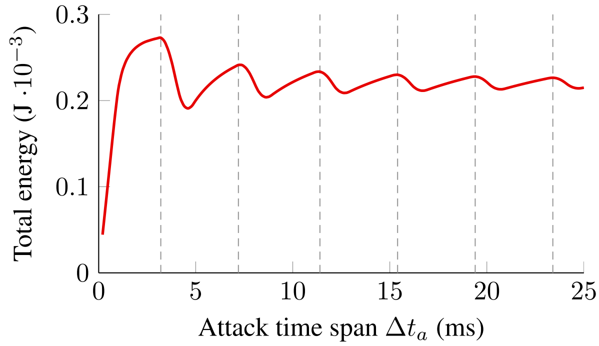


Figure 3. *Energy stored in the string for C_4 at release as a function of Δt_a.*

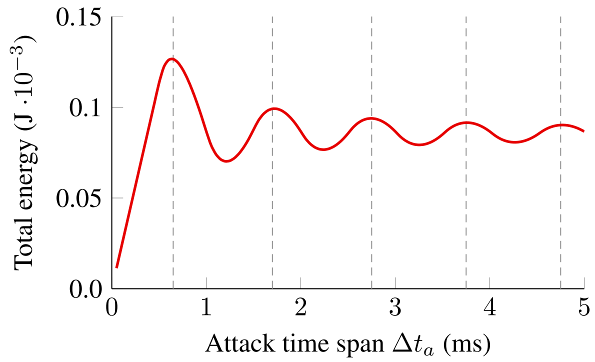


Figure 4. *Energy stored in the string for C_6 at release as a function of Δt_a.*

Figure 5 shows the work performed on C_4^*. Compared with Figure 3, it shows that the maxima and minima are more extreme for a string that is plucked in the middle, which is in line with the findings of Perng [4]. If the string is plucked near a fixed end, the first maximum is lower but there still is a wide range of Δt_a where there is an energy increase. (Compare Figures 3 and 5.)

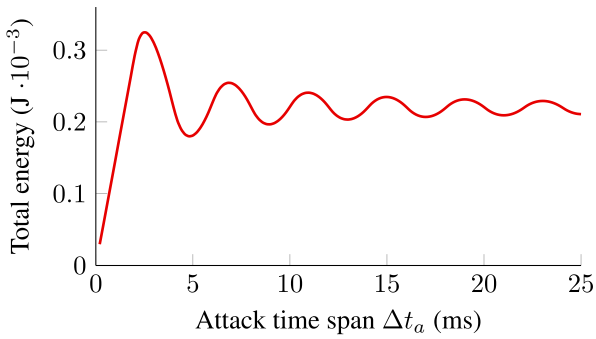


Figure 5. *Energy stored in the string for C_4^* (x_p = 350 mm) at release as a function of Δt_a .*

Figure 6 shows the kinetic energy for C_4 as a function of the attack time span Δt_a. This energy was computed as the sum of kinetic energies of all 1000 individual segments Δx

$$KE = \sum_{1 \leq i \leq 1000} 0.5 \mu \Delta x \left(\frac{\Delta u_i}{\Delta t} \right)^2 \tag{8}$$

Δu being the calculated change in string displacement in the time step Δt just before release. The kinetic energy is very low compared to the potential energy. (Note that the total energy in Figure 3 is in mJ whereas the kinetic energy in Figure 6 is in μJ.) This confirms the observation made in [3] that the amplitude of the string is as good an indicator of the loudness as is the energy.

5. THE ATTACK WAVE AND ITS REFLECTION

Figure 7 shows the shape of the string for several simulation times up to the moment of release for C_4, with an attack time span of $\Delta t_a = 3.2$ ms (for which the first maximum in energy is obtained). The time of first contact $t_0 = 0$, so the release is at $t = 3.2$ ms The straight grey lines from the maximum displacement at $x_p = 150$ mm to the fixed end at 700 mm represent the shape the string would have if it were lifted in a quasi steady-state ($\Delta t_a \to \infty$). The difference between the actual shape and the quasi steady-state shape gives an idea of the attack wave. The graph shows the initial stage of the creation of the attack wave at $t = 1.4$ ms where it also almost reaches the fixed end at the right. At $t = 2.0$ ms the reflection and the attack wave cancel out, and thereafter a crest of the reflection travels towards the plectrum.

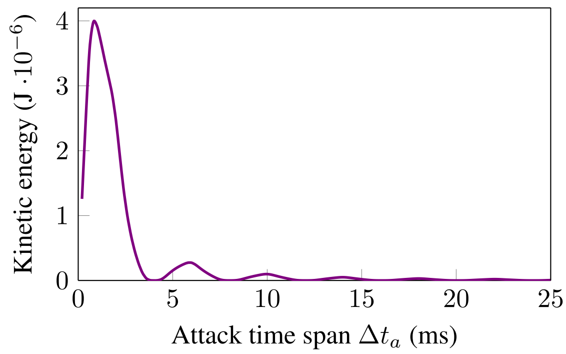


Figure 6. *Kinetic energy in the string for C_4 at the moment of release as a function of Δt_a.*

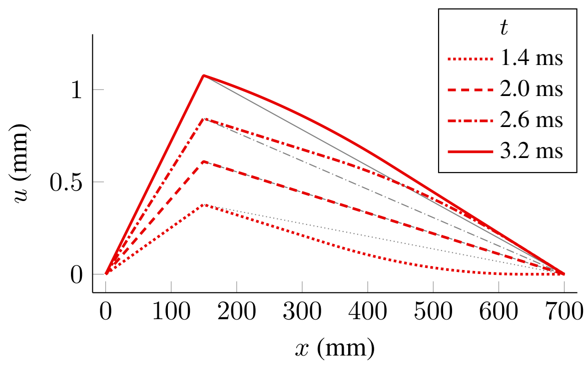


Figure 7. *String displacement for various simulation times, C_4 for Δt_a = 3.2 ms (maximum amplitude).*

Figure 8 shows a similar graph for an attack time span of $\Delta t_a = 4.6$ ms (for which the first minimum in energy is obtained). Again the formation of an attack wave and the reflection back towards the plectrum are apparent, but this time there is a trough of the reflection travelling towards the plectrum.

Figure 9 shows the displacement of the plucking point u_p over time for case C_4 for various attack time spans Δt_a. The first maximum displacement ($\Delta t_a = 3.2$ ms) is reached without the plucking point slowing down. In the case of the first minimum displacement ($\Delta t_a = 4.6$ ms) the plucking point is slowed down considerably before release, although it seems to gain a bit of speed shortly before release. The next maximum ($\Delta t_a = 7.2$ ms) has a phase where the plucking point is slowing down, after which it again gains speed to proceed up to release. The second minimum ($\Delta t_a = 8.6$ ms) is slowed down once, then gains speed and is then again slowed down in a similar fashion as

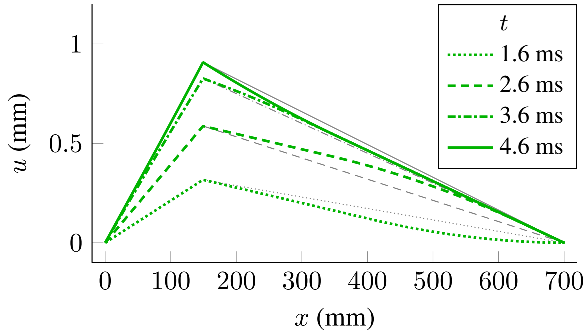


Figure 8. *String displacement for various simulation times, C_4 for Δt_a = 4.6 ms (minimum amplitude).*

for the first minimum before being released. The slowing down and speeding up is obviously the movement resulting from a prescribed force as in the current simulations.

Apparently, the first maximum occurs just before the reflection wave starts to exert a downward pressure on the plectrum. The first minimum occurs just after the reflection wave has exerted all possible downward force on the plectrum. Subsequent maxima and minima for larger attack time spans do the same for the second, third, etc. reflection of the attack wave. As the attack is slower for increasing Δt_a the amplitude of the attack wave is getting smaller and subsequently the maxima and minima are getting closer to the steady-state amplitude. Note that the plectrum is almost massless so that it will almost instantaneously yield to the force of the string. In fact, the natural frequency of the plectrum is about 2–8 kHz [8, 9] while the reflection wave has the frequency of the fundamental of the string. It therefore seems warranted to have omitted the behaviour of the plectrum in the model.

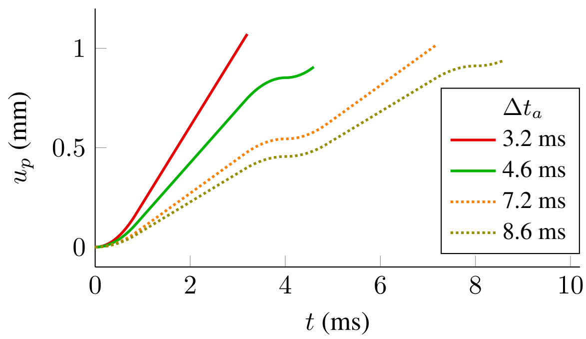


Figure 9. *Plucking point displacement until release, simulation C_4 for the first four extremes of energy.*

Comparing the shapes of the parts of the string left and right of the plucking point we can observe that the attack wave only gains a significant amplitude if the fixed end of the string is far enough away from the plectrum. The attack wave does not develop sufficiently on the shorter part of the string before it is reflected.

From the results for case C_4^* one can conclude that for a string that is plucked in the middle the attack wave runs symmetrically in both directions with the reflections from both fixed ends arriving at the plectrum at the same moment, thus reinforcing their action. Furthermore, although the plucking point is closer to the fixed end, the attack time spans where a significant increase of energy occurs are not very different from case C_4 where the plucking point is further away. One might have expected that (as the attack wave and its reflection have to travel a substantially shorter distance) the first maximum would occur sooner, but this does not seem to be the case. I haven't investigated this, but my conjecture would be that the attack must be executed more slowly to optimise the combination of the reflection waves from both the fixed ends.

6. TIMBRE

Figure 10 shows the signal of the transverse force that the string exerts on the soundboard for C_4, for attack time spans leading to the first maximum energy, the first minimum energy and the asymptotic steady-state energy. Although small vibrations in the longitudinal force also exist, the ones in the transverse force can be considered the main factor in determining the way the soundboard motion is driven by the string [10]. Given that the shapes of these waves differ from each other, one can conclude that there will be differences in the timbre between these three sounds. Small changes in timbre between a slow attack with a normal loudness and a fast attack with a higher loudness were indeed found in [5, 6].

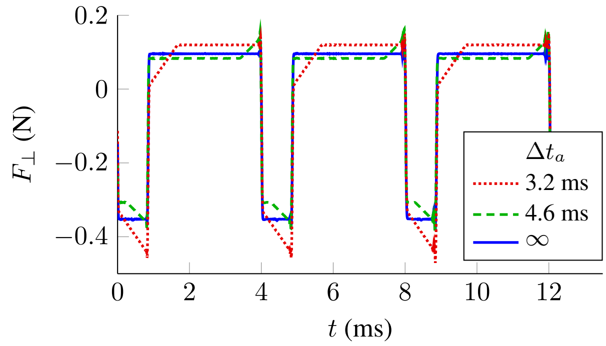


Figure 10. *Simulated time signals of the string force perpendicular to the bridge. The signals are shifted in time to enable an easy comparison. The small spikes at the beginning of a fall or a rise are artefacts resulting from numerical dispersion.*

7. LOUDNESS VARIATIONS IN HARPSICHORDS

7.1 General observations

The attack wave and its reflection lie at the heart of the variations in loudness. In the above this was analysed with a numerical model of the vibrating string. However, to be able to produce loudness variations in a real instrument the whole chain of the mechanical action between the finger and the plectrum lifting the string has to be taken into account.

In the discussion that follows I will focus on the first maximum energy that can be obtained when increasing the attack time span because of the following two reasons. Firstly, the experimental studies with real harpsichords only investigated a player's ability to produce higher volumes when striking the key faster. Secondly, subsequent minima and maxima are expected to be less pronounced than found in this study and therefore not of practical interest. The attack wave in the simulations was created by a prescribed force so there is no interaction with the displacement u_p of the plucking point. In reality the plectrum is reactive to the string movement and will cause secondary reflections that disperse the first reflection.

From the present study we can conclude that the possibility of producing an increased loudness of the plucked string with a faster key stroke in a harpsichord depends on two conditions:

1. The creation of an attack wave with sufficient amplitude;
2. The correct timing when lifting the string to allow the plectrum to rise further when the crest of reflections of the attack wave arrives back at the plucking point.

The first condition is evident from the fact that an increasing attack time span produces progressively lower maxima, asymptotically converging to a steady-state loudness. The second condition follows from the analysis of the reflection of the attack wave in Section 5.

Still another condition that could possibly be added is the shape of the attack wave. This was not addressed in the present study, but it can be imagined that different patterns of an increase in the force on the string might change the effectiveness of the action of the reflection wave back on the plectrum.

In the following sections, using the results of the simulations C_2, C_4 and C_6, I will first consider the second condition in the context of playing a harpsichord. The result is what could be called a loudness map. Next I will draw up a list of factors pertaining to the construction and regulation of an instrument that can conceivably influence the opportunities a player has for creating the first and/or the second condition.

7.2 The loudness map

In Figure 11 I have drawn the loudness map for the lower manual of the Hemsch instrument that served as a reference for the simulations. Along the horizontal axis is the keyboard range and along the vertical axis are the attack time spans Δt_a. The orange area shows the attack time spans required to produce a 10% increase in the work W_s compared to a steady-state pluck. (It is thus essentially a contour plot.)

The striped blue area represents key dip velocities that correspond to normal playing, which are smaller than 0.1 m/s [4, 11]. The pink band at the bottom represents the maximum key dip velocity that can practically be obtained during a performance by striking with more force [3, 4, 8]. These key velocities were translated into attack time spans assuming: **a)** a lever ratio of the key around the balance point (front length divided by rear length) of 0.77; **b)** the maximum amplitudes listed in Table 2; **c)** a slight deviation of the string away from the jack so that the string deviation at the plucking point is about 10 degrees off from the vertical [7]; and **d)** that the key dip velocity is constant.

The grey dotted area is where the simulations indicate that the energy of the string should drop to almost zero. This was however not confirmed in experiments [3]. Trying this myself on an instrument I was also not able to produce any sound that was significantly softer with a very fast attack. At present, it therefore remains uncertain what exactly the resulting energy is in this area.

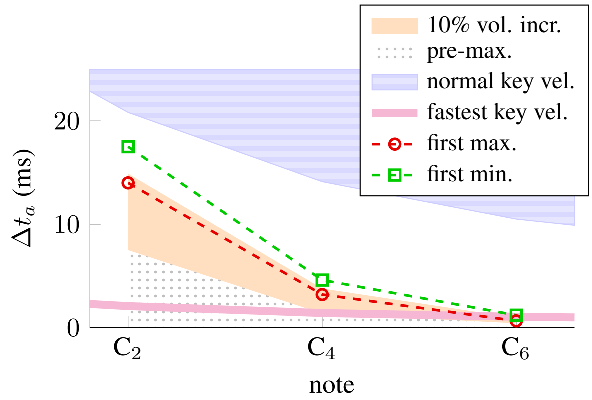


Figure 11. *Loudness map for the lower manual of a Hemsch harpsichord.*

As a reference I also included the lines representing the first maximum and the first minimum. The closer the attack time span gets to the red dashed line (the first maximum), the louder the resulting volume. (If the timing is not accurate enough, one may actually end up with a somewhat smaller volume, indicated by the green dashed line, although as mentioned before this minimum is expected to be of no practical importance. Also, a performer may still benefit from the knock effect [5].)

The loudness map shows that normal playing (the striped light blue area) takes place entirely in the region where there are no significant loudness variations, in line with what is observed in practice. Producing a louder tone is well within a player's reach for lower notes. Here the required attack time span is well above the minimum that can be practically achieved and there is also a wider range of valid time spans. Going up towards the treble the attack time span not only gets closer to what is achievable, but the timing also needs to be increasingly accurate. In the very high treble, the attack time span may even be too short to obtain loudness variations under practical conditions.

In the case under consideration (the lower manual of a Hemsch), the relative plucking point $p_r = x_p/L$ is smaller in the bass than in the treble and the maximum relative increase in loudness would therefore also be less than in the treble. A plucking point that is more towards the middle ($p_r \rightarrow 0.5$) would increase the maximum obtainable loudness and also shift the red dashed line (the first maximum) down towards smaller attack time spans. The area of 10% loudness increase is not expected to change much for a plucking point more towards the middle. (Compare Figures 3 and 5.)

7.3 Influence of construction and regulation

In order to be able to draw the loudness map I had to resort to several assumptions about the configuration of the mechanical action of the instrument. Indeed, the exact layout of the loudness map depends on details of the construction and regulation. These details can influence the timing that is achieved with a certain force pressing down the key and/or influence the shape and thus the effectiveness of the attack wave.

What follows is a list of factors that can conceivably influence the potential of producing loudness effects on a particular instrument. This list contains some speculation and may not be exhaustive, but it can help raise awareness of the complexity involved in producing louder tones on a harpsichord. (See Figure 1 for a schematic diagram of a harpsichord mechanism.)

1. The construction of the key:
   a) The weight of the key and its weight distribution (together expressed in the moment of inertia) as well as the weight of the jacks. A higher moment of inertia and heavier jacks will mean that it takes more force to accelerate the key and jacks to give sufficient velocity to obtain the maximum amplitude. The effects can be quite pronounced even for small amounts of extra mass, especially for shorter attack time spans [8].
   b) The position of the balance point, which defines the ratio of the downward velocity at the front of the key to the upward velocity of the jack.
   c) The stiffness of the key lever. The lever may bend slightly, especially the longer levers that are found in the lower manual of double manual instruments, and particularly when there are higher forces involved [8]. This would make the translation between the key dip velocity and the jack velocity non-linear and thus influence the shape and magnitude of the attack wave.

2. The transmission of the movement from the key to the jack:
   a) The elasticity of the cloth on the back of the key where the jack sits.
   b) The surface area of the bottom end of the jack. In historical harpsichords the bottoms of jacks vary from a flat wooden surface over the entire cross section of the jack or a part of it (as in "dogleg" jacks or Taskin's "peau de buffle") to a rounded end (as for example in the 1754 Hemsch currently in the Bayrisches Nationalmuseum in Munich). In modern instruments sometimes small adjustment screws are inserted at the bottom of the jack which leads to a small contact area.

3. The design of the jacks. Historically jacks were made of wood and the plectrum was inserted in a wooden component, called the tongue, which can rotate around an axle but is kept in place by a spring, normally a hog's bristle or sometimes a thin brass strip. The design of the mechanism is such that during the lifting of the string the tongue is completely checked by the body of the jack which guarantees a very stable position, resulting in a well defined moment of release of the string. Some modern builders use plastic jacks. In one common type of such jacks the tongue, axle and spring are all integrated into one element, a tongue with a special design. The action of this jack is, however, not as precise as a historical wooden jack because when applying pressure on the plectrum this special tongue yields somewhat. Thus the plectrum does not only bend, it also rotates a fraction at its base. (Unfortunately experiments are sometimes carried out with such jacks, which may make certain results questionable.)

4. The length of the plectrum, its stiffness and the stiffness distribution along its length. These factors determine the deflection at the tip of the plectrum when the string is released [8]. This in turn determines the distance the jack has to travel from the first contact between string and plectrum to the release, and thus the key dip velocity needed for a certain attack time span Δt_a.

5. The material of the plectrum. The traditional material in past centuries was bird quill. In the second half of the 20th century a plastic substitute – polyacetal, commonly referred to with the Dupont brand name Delrin® – has become popular. This is a polymer with a viscoelastic behaviour that is essentially of a Kelvin-Voigt type with a relatively low amount of damping. Bird quill seems to exhibit more the behaviour of a Maxwell material and may have a substantially different deflection over time when absorbing the shock of the impact between plectrum and string [8]. In this context I would venture to say that the fact that it has not been possible to experimentally obtain vanishing energies in the string for small Δt_a (the dotted area in Figure 11) may be due to the visco-elastic behaviour of plectrum materials in combination with an upper boundary of achievable jack velocities. The use of leather is a case on its own, because the friction between leather and metal is substantially higher than between polyacetal or quill and metal. It also has a lower hardness, so the string may bury itself somewhat in the surface.

6. The string tension. Thicker strings need smaller amplitudes to have the same energy as thinner strings, so they require a lower jack velocity and thus make it easier to achieve the key velocity required for a louder tone.

8. CONCLUSIONS

Using a numerical model, the attack wave of the plectrum hitting, lifting and releasing a harpsichord string was analysed. The reflections of the attack wave off the fixed ends of the string can cause an increase in the loudness of the resulting sound, confirming the suggestions from other studies. The conditions that allow for an increase in loudness in an individual string depend on the correct timing of the arrival of the reflection wave and its magnitude. If the crest of the reflection reaches the location of the plectrum just before the release of the string, the plectrum will be able to lift the string further up before release.

The timing is mainly governed by the frequency of the note and thus varies over the range of the keyboard. Key dip velocities for normal playing generate timings where there are no practical loudness variations, but striking a key faster can lead to an increase in loudness provided the correct timing is achieved. This is easier for lower notes because the necessary key dip velocity is well below the maximum velocity that can be practically achieved. Also, there is a relatively wide range of velocities that produce an increase in loudness so the exact value does not need to be very accurate. Going up towards the treble, the attack time span needs to be progressively smaller. Not only does it get closer to the limits of what is practically achievable, but the timing needs to be increasingly accurate as well.

Other influencing factors are the details of the mechanical action between the movement of the finger and the lifting of the string. These details are determined by the construction and regulation of a harpsichord. Therefore, individual instruments, and individual registers within an instrument, may provide better or worse conditions for actually achieving variations in loudness.

It was also shown that the amplitude of the plucking point at release is a good indicator of the loudness, because the kinetic energy contained in the string at the moment of release is very small compared to the potential energy. In addition, timbre variations can be expected for notes that have been produced with higher key dip speeds.

9. REFERENCES

[1] D. E. Hall, "String excitation: piano, harpsichord and clavichord," in *Proceedings of Stockholm Music Acoustics Conference (SMAC)*, 1993, pp. 309–314.

[2] D. Griffel, "The dynamics of plucking," *Journal of Sound and Vibration*, vol. 175, 1994.

[3] N. Giordano and J. P. Winans II, "Plucked strings and the harpsichord," *Journal of Sound and Vibration*, vol. 224, 1999.

[4] C.-Y. J. Perng, "Physical modeling of the harpsichord plectrum-string interaction," Ph.D. dissertation, Stanford University, 2012.

[5] H. Penttinen, "On the dynamics of the harpsichord and its synthesis," in *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)*, 2006, pp. 115–120.

[6] J. MacRitchie and G. Nuti, "Using historical accounts of harpsichord touch to empirically investigate the production and perception of dynamics on the 1788 Taskin," *Frontiers in Psychology*, vol. 6, 2015.

[7] A. Paté, J.-L. Le Carrou, A. Givois, and A. Roy, "Influence of plectrum shape and jack velocity on the sound of the harpsichord: An experimental study," *The Journal of the Acoustical Society of America*, vol. 141, 2017.

[8] G. Karman, "Ruminations on harpsichord building," Unpublished manuscript, due for publication in 2024.

[9] T. Gäumann, "The pretransient of the harpsichord sound. I. The vertical movement of the plectrum," in *Proceedings of Stockholm Music Acoustics Conference (SMAC)*, 2003, pp. 163–166.

[10] C. Valette and C. Cuesta, *Mécanique de la corde vibrante*. Paris: Hermes, 1993.

[11] D. Chadefaux, J.-L. Le Carrou, S. L. Conte, and M. Castellengo, "Analysis of the harpsichord plectrum-string interaction," in *Proceedings of Stockholm Music Acoustics Conference (SMAC)*, 2013, pp. 155–160.

A Fast Algorithm for the Inversion of the Biharmonic in Plate Dynamics Applications

Zehao Wang
Department of Music
University of California, San Diego
La Jolla, CA, USA
zehaowang@ucsd.edu

Miller Puckette
Department of Music
University of California, San Diego
La Jolla, CA, USA
msp@ucsd.edu

ABSTRACT

In this paper, we present a numerical method for solving the biharmonic equation using finite difference methods, which can be used for fast acoustic simulation with nonlinear plate dynamics. With the simply supported boundary condition, the linear system could be regarded as a composition of two Poisson's equations, and these Poisson's equations are solved by the Thomas algorithm for a series of tridiagonal systems after transpositions and linear transformations for vectors in the systems and all nonempty blocks of the Laplacian matrix. We also point out that the eigendecomposition used for these linear transformations has a closed-form formula, which is easy to be pre-computed and also space-saving. Furthermore, since this solver is computed block by block and does not need sparse matrix operations, this method is good for single instruction multiple data (SIMD) parallelization using advanced vector extensions (AVX) intrinsics on central processing units (CPUs), which makes it possible to execute at high speeds for real-time music applications. We also show that this solver for the simply supported boundary condition can also be easily adapted for other boundary conditions using Woodbury matrix identity with a little extra complexity. Numerical experiments show that the C++ implementation of this method is faster than decomposition-based solvers (like LU or Cholesky decomposition) of some well-known C++ libraries at the scale of applications in the field of musical acoustics.

1. INTRODUCTION

Physical modeling methods have a long-established history in simulating musical instruments. This involves representing a particular musical instrument using a system of differential equations, which can be solved using various numerical techniques such as finite-difference, finite-element, and finite-volume methods. The application of physical modeling extends to both musical acoustics, facilitating an examination of the intricate dynamics of musical instruments, and sound synthesis. Of particular importance are the strongly nonlinear effects that underlie the behavior of numerous musical instruments, which present significant challenges in terms of algorithmic design and computation cost.

The simulation of nonlinear plate dynamics problems, such as that of von Kármán [1, 2], typically requires the inversion of the biharmonic operator, which is a computational bottleneck, which guides us to find a fast solver for the linear system with the biharmonic operator.

In the realm of scientific computing and computational mathematics, a range of sparse matrix solvers have been developed to address numerical partial differential equations (PDEs) using methods like fast Fourier transform (FFT) [3], matrix decomposition, or iterative approaches [4]. However, the computational concerns that these methods address are generally related to scalability, which differs significantly from the needs of acoustic simulation, particularly in fast simulation scenarios. In general, algorithms for fast acoustic simulation like sound synthesis should be suitable for low-level SIMD parallelization on CPUs such as advanced vector extensions (AVX) intrinsics, since such optimization methods show great efficiency in the application of fast musical acoustic simulation scenarios using finite-difference schemes [5]. While methods that exploit the structure of sparse matrices, such as FFT-based or cyclic-reduction-based methods [3, 6], can be effective, they may not perform optimally for the scale of musical instrument simulation, since the scale of problems involved in this application is relatively small. For instance, some cyclic-reduction-based methods stop doing cyclic reductions when the matrix size is around 3×3 to 7×7 and directly solve it instead. Nonetheless, in many musical instrument simulation problems, the grid size required for acceptable sound quality ranges from 15×15 to 40×40, meaning that only a limited amount of time for cyclic reductions will be executed, and these operations may take extra time, which is a significant concern at this scale and cannot be optimized much by low-level SIMD parallelization.

In order to accelerate the performance of specific problems of modeling nonlinear plate dynamics, i.e., the fast inverse of the biharmonic operator, we propose a method adapted from [6] to solve the linear system that appeared in most of the schemes for this model with a high speed. After a series of transpositions and linear transformations derived from the closed-form eigendecomposition of a given matrix, the original system could be solved by applying

Thomas algorithm [7] which only requires linear time cost to diagonal blocks. Optimization techniques for C++ implementations like loop unrolling or SIMD parallelization using AVX intrinsics which are compatible with different platforms are also proposed to achieve high speed. Numerical results show that the C++ implementation of this solver could be optimized a lot by those techniques, and its performance is much better than several widely-used solvers, and the timing results show the possibility of fast simulation of plate models with nonlinear plate dynamics. A real-time algorithm for the solution of the von Kármán system for real-time synthesis of gong-like sounds is under development [8].

2. PRELIMINARIES

2.1 The von Kármán plate model

For modeling the nonlinear vibration of plates at moderate amplitudes, the following von-Kármán equation is commonly used:

$$\rho H u_{tt} = -D\Delta\Delta u + \mathcal{L}(\Phi, u), \tag{1a}$$

$$\Delta\Delta\Phi = -\frac{EH}{2}\mathcal{L}(u, u), \tag{1b}$$

where

$$\mathcal{L}(\alpha, \beta) = \alpha_{xx}\beta_{yy} + \alpha_{yy}\beta_{xx} - 2\alpha_{xy}\beta_{xy}, \tag{2}$$

$\Phi(x, y, t)$ is the airy stress function.

The followings are two sets of boundary conditions (clamped and simply supported, respectively) over the boundary ∂U of the domain U:

$$u = \frac{\partial}{\partial \mathbf{n}}u = 0, \quad \Phi = \frac{\partial}{\partial \mathbf{n}}\Phi = 0 \qquad \text{clamped}, \tag{3a}$$

$$u = \Delta_{\mathrm{n}} u = 0, \quad \Phi = \Delta_{\mathrm{n}}\Phi = 0 \qquad \text{simply supported}, \tag{3b}$$

where $\frac{\partial}{\partial \mathbf{n}}$ and Δ_{n} denote the first-order and second-order scalar derivative in the normal direction of the boundary ∂U.

This paper is not concerned with the numerical solution of the von Kármán equations, but rather with the problem of the inversion of the biharmonic operator that appears in (1b).

2.2 Grid functions and finite difference operators

Assume the domain of interest U is a rectangular domain with side lengths $L_x \times L_y$, and its discretization is $N_x \times N_y$ with grid spacing $h_x = h_y = h$, where $h_x = L_x/N_x$ and $h_y = L_y/N_y$.

For a given grid function $u_{l,m}$, define the following spatial difference operator to approximate the derivative operators:

$$\delta_{x\pm} \triangleq \pm\frac{1}{h}\left(e_{x\pm} - 1\right) \approx \frac{\partial}{\partial x}, \delta_{y\pm} \triangleq \pm\frac{1}{h}\left(e_{y\pm} - 1\right) \approx \frac{\partial}{\partial y},$$

where $e_{x\pm}u^n_{l,m} = u^n_{l\pm1,m}$ and $e_{y\pm}u^n_{l,m} = u^n_{l,m\pm1}$. Then we define centered second derivative approximations as follows,

$$\delta_{xx} = \delta_{x+}\delta_{x-} \approx \frac{\partial^2}{\partial x^2}, \quad \delta_{yy} = \delta_{y+}\delta_{y-} \approx \frac{\partial^2}{\partial y^2}$$

The Laplacian and biharmonic operators may then be approximated as

$$\delta_{\Delta\boxplus} = \delta_{xx} + \delta_{yy} \approx \Delta, \quad \delta_{\Delta\boxplus,\Delta\boxplus} \triangleq \delta_{\Delta\boxplus}\delta_{\Delta\boxplus} \approx \Delta\Delta.$$

With simply support boundary condition, we can also write them in matrix form as

$$\mathbf{D}_{\boldsymbol{\Delta}} = \mathbf{L}/h^2, \quad \mathbf{D}_{\boldsymbol{\Delta\Delta}} = \mathbf{D}_{\boldsymbol{\Delta}}\mathbf{D}_{\boldsymbol{\Delta}} = \mathbf{L}^2/h^4 = \mathbf{B}/h^4,$$

here

$$\mathbf{L} = \begin{bmatrix} \mathbf{A} & \mathbf{I} & & & 0 \\ \mathbf{I} & \mathbf{A} & \mathbf{I} & & \\ & \cdots & \cdots & & \\ & & \cdots & \cdots & \\ & & \mathbf{I} & \mathbf{A} & \mathbf{I} \\ 0 & & & \mathbf{I} & \mathbf{A} \end{bmatrix} \in \mathbb{R}^{NN\times NN}, \tag{4}$$

where $NN = (N_y-1)(N_x-1)$, L is the discrete Laplacian operator, $I \in \mathbb{R}^{(N_y-1)\times(N_y-1)}$ is the identity matrix, and

$$\mathbf{A} = \begin{bmatrix} -4 & 1 & & & 0 \\ 1 & -4 & 1 & & \\ & \cdots & \cdots & & \\ & & \cdots & \cdots & \\ & & 1 & -4 & 1 \\ 0 & & & 1 & -4 \end{bmatrix} \in \mathbb{R}^{(N_y-1)\times(N_y-1)}.$$

2.3 FDTD schemes

To numerically solve the system (1), a number of schemes could be used [2, 9–11]. Here we only focus on the major computational bottleneck of FDTD schemes which is to solve discrete Φ from (1b), which requires us to find the solution of a linear system. In general, the linear system should have the following formula,

$$\delta^4[\Phi] = d, \tag{5}$$

where δ^4 is the discrete counterpart of $\Delta\Delta$, and d is the discrete vector of the right-hand side of Eq. (1b) derived by a given FDTD scheme. With the simply supported boundary condition which is often used for sound synthesis [2, 10], the form of δ^4 we consider here is $\mathbf{D}_{\boldsymbol{\Delta\Delta}}$. Thus, the linear system to be solved is equivalent to

$$\mathbf{B}[\Phi] = h^4\mathbf{D}_{\boldsymbol{\Delta\Delta}}[\Phi] = h^4 d. \tag{6}$$

2.4 A decomposition of A

Notice that $\mathbf{A}$ has a decomposition $\mathbf{Q}^*\mathbf{V}\mathbf{Q}$, where $\mathbf{Q}$ is a unitary matrix [1] and $\mathbf{V}$ is a diagonal matrix. $\mathbf{Q}$ and $\mathbf{V}$ have the following closed-form formulas,

$$\mathbf{Q}_{kj} = \sqrt{\frac{2}{N_y}}\sin\left(\frac{kj\pi}{N_y}\right), \tag{7}$$

[1] $\mathbf{Q} = \mathbf{Q}^*$ and $\mathbf{Q}\mathbf{Q}^* = \mathbf{I}$, actually here we have $\mathbf{Q}* = \mathbf{Q}^T$.

$$\mathbf{V}_{kk} = 2\cos\left(\frac{k\pi}{N_y}\right) - 4, \tag{8}$$

where $1 \leq k, j \leq N_y - 1$. To prove this decomposition, we only need to prove the following lemma:

Lemma 2.1. *$\mathbf{A}$'s (N_y-1) distinct eigenvalues are $\mathbf{V}_{11}, \mathbf{V}_{22}, \ldots, \mathbf{V}_{(N_y-1)(N_y-1)}$, and $\mathbf{Q}_{k*}$ is the unit eigenvector of $\mathbf{A}$ with respect to $\mathbf{V}_{kk}$, $1 \leq k \leq N_y - 1$.*

The proof is shown in Section 7.

2.5 Thomas algorithm for tridiagonal systems

The Thomas algorithm [7] for tridiagonal systems (9) is shown in Algorithm 1.

$$\mathbf{M}x = \mathbf{M}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = y, \tag{9}$$

where

$$\mathbf{M} = \begin{bmatrix} b_1 & c_1 & & & 0 \\ a_2 & b_2 & c_2 & & \\ & \cdots & \cdots & & \\ & & \cdots & \cdots & \\ & & a_{n-1} & b_{n-1} & c_{n-1} \\ 0 & & & a_n & b_n \end{bmatrix}_{n\times n},$$

Algorithm 1 Thomas algorithm

Input: $y = [y_1, y_2, \ldots, y_n]^T, b = [b_1, b_2, \ldots, b_n]^T \in \mathbb{R}^n$,
$a = [a_2, a_3, \ldots, a_n]^T, c = [c_1, c_2, \ldots, c_{n-1}]^T \in \mathbb{R}^{n-1}$

Output: $x = [x_1, x_2, \ldots, x_n]^T \in \mathbb{R}^n$

```
function ThomasAlgorithm(a, b, c, y)
    Forward elimination:
    for i = 2 to n do
        w ← a_{i-1}/b_{i-1}
        b_i ← b_i − w c_{i-1}
        y_i ← y_i − w y_{i-1}
    end for
    x_n ← y_n/b_n
    Backward substitution:
    for i = n − 1 to 1 by -1 do
        x_i ← (y_i − c_i x_{i+1})/b_i
    end for
    return x
end function
```

A simple sufficient condition to ensure the stability of Algorithm 1 is diagonally dominant (either by row or column) [4] [2], which means $|b_i| \geq |a_i| + |c_i|,\ i = 1, 2, \ldots n$ for system (9) [3].

[2] There are other sufficient conditions for the stability of such systems.
[3] Assume $a_1 = c_n = 0$.

3. THE BIHARMONIC SOLVER

In short, the linear system (6) could be solved by the Thomas algorithm for a series of tridiagonal systems after transpositions and linear transformations for vectors in the systems and all non-empty blocks of the Laplacian matrix. In this section, we will develop details of the biharmonic solver.

Since the discrete biharmonic system with simply supported conditions can be regarded as a composition of two discrete Laplacian systems,

$$b = \mathbf{B}x = \mathbf{L}\mathbf{L}x, \tag{10}$$

the system could be solved by applying the above Laplacian solver twice,

$$\begin{cases} \mathbf{L}v = b \\ \mathbf{L}x = v \end{cases}, \tag{11}$$

which means we first solve v from the first equation, and solve u from the second equation. Thus, we first introduce the solver for discrete Laplacian systems using this decomposition and the Thomas algorithm [7] [4] from [6] first.

Let

$$\tilde{x} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1(N_x-1)} \\ x_{21} & x_{22} & \cdots & x_{2(N_x-1)} \\ \vdots & \vdots & \vdots & \vdots \\ x_{(N_y-1)1} & x_{(N_y-1)2} & \cdots & x_{(N_y-1)(N_x-1)} \end{bmatrix}$$

be unknown on the grids of the discrete rectangular plate area, and $x = [(\tilde{x}_{*1})^T, (\tilde{x}_{*2})^T, \ldots, (\tilde{x}_{*(N_x-1)})^T]^T$ is the flattened vector of $\tilde{x}$ by column. Thus, the discrete Laplacian system we need to solve is as follows,

$$\mathbf{L}x = b, \tag{12}$$

where $\tilde{b}$ is the known on the grids of the discrete rectangular plate area and b is the flattened vector of $\tilde{b}$ by column. Consider the block structure of L in (4), we have the following equivalent system

$$\begin{array}{ccccccc} & & \mathbf{A}\tilde{x}_{*1} & + & \tilde{x}_{*2} & = & \tilde{b}_{*1}, \\ \tilde{x}_{*(j-1)} & + & \mathbf{A}\tilde{x}_{*j} & + & \tilde{x}_{*(j+1)} & = & \tilde{b}_{*j}, \\ \tilde{x}_{*(N_x-2)} & + & \mathbf{A}\tilde{x}_{*(N_x-1)} & & & = & \tilde{b}_{*(N_x-1)}, \end{array} \tag{13}$$

for $j = 2, 3, \ldots, N_x - 2$.

Consider the transformation for $\bar{x}_{*j} = \mathbf{Q}^*\tilde{x}_{*j}$ and $\bar{b}_{*j} = \mathbf{Q}^*\tilde{b}_{*j}$ [5], and multiply Q^* on both sides of each equation in (13), we have the following equivalent system

$$\begin{array}{ccccccc} & & \mathbf{V}\bar{x}_{*1} & + & \bar{x}_{*2} & = & \bar{b}_{*1}, \\ \bar{x}_{*(j-1)} & + & \mathbf{V}\bar{x}_{*j} & + & \bar{x}_{*(j+1)} & = & \bar{b}_{*j}, \\ \bar{x}_{*(N_x-2)} & + & \mathbf{V}\bar{x}_{*(N_x-1)} & & & = & \bar{b}_{*(N_x-1)}, \end{array} \tag{14}$$

for $j = 2, 3, \ldots, N_x - 2$.

[4] A direct method based on LU decomposition for solving tridiagonal systems with time complexity of $O(n)$, where $n \times n$ is the size of the matrices.
[5] where $\bar{x} = [\bar{x}_{*1}, \bar{x}_{*2}, \ldots, \bar{x}_{*(N_x-1)}]$ and $\bar{b} = [\bar{b}_{*1}, \bar{b}_{*2}, \ldots, \bar{b}_{*(N_x-1)}]$.

Consider each entry in each equation of (14), the system could be rewritten for $k = 1, 2, \dots, N_y - 1$,

$$\begin{array}{ccccccc} & & \mathbf{V}_{kk}\bar{x}_{k1} & + & \bar{x}_{k2} & = & \bar{b}_{k1}, \\ \bar{x}_{k(j-1)} & + & \mathbf{V}_{kk}\bar{x}_{kj} & + & \bar{x}_{k(j+1)} & = & \bar{b}_{kj}, \\ \bar{x}_{k(N_x-2)} & + & \mathbf{V}_{kk}\bar{x}_{k(N_x-1)} & & & = & \bar{b}_{k(N_x-1)}, \end{array} \tag{15}$$

for $j = 2, 3, \dots, N_x - 2$.

Now denote

$$\mathbf{\Gamma}_k = \begin{bmatrix} \mathbf{V}_{kk} & 1 & & & & \\ 1 & \mathbf{V}_{kk} & 1 & & & \\ & \cdots & \cdots & & & \\ & & \cdots & \cdots & & \\ & & 1 & \mathbf{V}_{kk} & 1 \\ & & & 1 & \mathbf{V}_{kk} \end{bmatrix}_{(N_x-1)\times(N_x-1)}, \tag{16}$$

for $k = 1, 2, \dots, N_y - 1$, $\hat{x} = \bar{x}^T$, and $\hat{b} = \bar{b}^T$, where $\hat{x}_{*k} = [\bar{x}_{k1}, \bar{x}_{k2}, \dots, \bar{x}_{k(N_x-1)}]^T$ and $\hat{b}_{*k} = [\bar{b}_{k1}, \bar{b}_{k2}, \dots, \bar{b}_{k(N_x-1)}]^T$. Then we have the following system,

$$\mathbf{\Gamma}_k \hat{x}_{*k} = \hat{b}_{*k}, \quad k = 1, 2, \dots, N_y - 1, \tag{17}$$

which is equivalent to (13), (14), and (15). Consider the tridiagonal systems in (17), we can easily show that

$$|\mathbf{V}_{kk}| = |2\cos(\frac{k\pi}{N_y}) - 4| \geq 4 - 2|\cos(\frac{k\pi}{N_y})| \geq |1|,$$

for $k = 1, 2, \dots N_y - 1$, which means all Γ_ks are diagonally dominant so that Algorithm 1 is stable for systems in (17). Since Γ_ks have simple structures, we can simplify Algorithm 1 to Algorithm 2.

Algorithm 2 Thomas algorithm simplified for $\mathbf{\Gamma}_k$s

Input: $\lambda(= V_{kk}), y = [y_1, y_2, \dots, y_n]^T \in \mathbb{R}^n$
Output: $x = [x_1, x_2, \dots, x_n]^T \in \mathbb{R}^n$

```
function SIMPLIFIEDTHOMASALGORITHM(λ, y)
    Initialize an empty vector q = [q_1, q_2, ..., q_n]^T ∈ ℝ^n
    q_1 ← λ
    Forward elimination:
    for i = 2 to n do
        w ← 1/q_{i-1}
        q_i ← λ − w
        y_i ← y_i − w y_{i-1}
    end for
    x_n ← y_n/q_n
    Backward substitution:
    for i = n − 1 to 1 by -1 do
        x_i ← (y_i − x_{i+1})/q_i
    end for
    return x
end function
```

Notice that after we solve the above system, we can easily transform $\bar{x}$ (which equals to $\hat{x}^T$) back into $\tilde{x}$ by $\tilde{x}_{*k} = \mathbf{Q}\bar{x}_{*k}$.

As we can see above, the original discrete Laplacian system (13) could be transformed into a series of tridiagonal systems (17), and those tridiagonal systems could be solved by the Thomas algorithm (shown in Algorithm 1 and 2).

As we mentioned before, to solve the biharmonic system (10), we only need to solve two Laplacian systems (11). However, notice that $\mathbf{QQ}^* = \mathbf{I}$, there are two linear transformations at the beginning ($\mathbf{Q}^*$) and the end of the algorithm ($\mathbf{Q}$), and the solution of the first Laplacian system should be the right-handed side of the second Laplacian system, which means we can simply discard the last transformation in the first solver and the beginning transformation of the second solver without changing the result for the whole biharmonic system. A brief diagram is shown in Fig. 1, and the pseudo-code for this solver is given in Algorithm 3.

3.1 Other boundary conditions

For other boundary conditions, like the clamped boundary condition or the free boundary condition, the matrix δ^4 in Eq. (5) could be written as $B + UV$, where $\mathbf{B} = \mathbf{L}^2 \in \mathbb{R}^{NN\times NN}$ is the discrete biharmonic operator we derived before, $\mathbf{U} \in \mathbb{R}^{NN\times m}$ and $V \in \mathbb{R}^{m\times NN}$. Actually, for most boundary conditions, $\mathbf{V} = \mathbf{U}^T$ and $\mathbf{U}$ is a rank-m sparse matrix with cNN nonzero entries, where $c = 1$ for the clamped boundary condition.

Therefore, one can use the following formula adapted from Woodbury matrix identity [12],

$$(\mathbf{B} + \mathbf{UV})^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{U}\left(\mathbf{I} + \mathbf{VB}^{-1}\mathbf{U}\right)^{-1}\mathbf{VB}^{-1}, \tag{18}$$

which means we need to precompute and store several matrices like $\mathbf{B}^{-1}\mathbf{U}$ using the biharmonic solver we described before, and do some extra sparse matrix-vector multiplication (like vectors multiplied by $\mathbf{U}$) and dense matrix-vector multiplication (like vectors multiplied by $\mathbf{B}^{-1}\mathbf{U}$) at each time step.

4. IMPLEMENTATION

4.1 Implementation and optimization of the biharmonic solver

Consider the linear transformation stage and the Thomas algorithm stage, each column of $\tilde{x}$ and $\hat{x}$ is individual, so the optimization technique is to unroll or parallelize using AVX intrinsics every for-loop in Algorithm 3, and use AVX's fused operations like fused multiply-add (fmadd) instead of two separate operations if supported. However, the optimization using AVX is a little complicated since all matrices that will be parallelized need to be stored by some specific orders by column or row [6]. A brief demonstration of these optimization techniques is shown in Fig. 2. Although the transpose operations introduce extra complexity, numerical results from the next section show that the overall performance of the AVX version is better than the plain version and loop-unrolling version, and the loop-unrolling version which can be used if AVX is not compatible with the hardware is a little slower than the AVX version but faster than the plain version.

[6] For example, if the index m that will be parallelized indicates m-th column, the array should be flattened by row, and vice versa.

$$\boxed{\mathbf{L}v = b} \;\Rightarrow\; b \xrightarrow{\text{flatten}} \tilde{b} \;\Rightarrow\; \hat{b}_{j*} = (\bar{b}_{*j})^T = (\mathbf{Q}^* \tilde{b}_{*j})^T \;\Rightarrow\; \hat{v}_{*k} = \mathbf{\Gamma}_k^{-1} \hat{b}_{*k}$$

$$\cancel{\Rightarrow\; \tilde{v}_{*j} = \mathbf{Q}\bar{v}_{*j} = \mathbf{Q}(\hat{v}_{j*})^T \;\Rightarrow\; \tilde{v} \xrightarrow{\text{reshape}} v \;\Rightarrow\; \boxed{\mathbf{L}x = v} \;\Rightarrow\; v \xrightarrow{\text{flatten}} \tilde{v} \;\Rightarrow\; \hat{v}_{j*} = (\bar{v}_{*j})^T = (\mathbf{Q}^* \tilde{v}_{*j})^T}$$

$$\Rightarrow\; \hat{x}_{*k} = \mathbf{\Gamma}_k^{-1} \hat{v}_{*k} \;\Rightarrow\; \tilde{x}_{*j} = \mathbf{Q}\bar{x}_{*j} = \mathbf{Q}(\hat{x}_{j*})^T \;\Rightarrow\; \tilde{x} \xrightarrow{\text{reshape}} x.$$

Figure 1. A diagram of the biharmonic solver. The strikethrough across the second line means that these operations can be discarded because they cancel each other out.

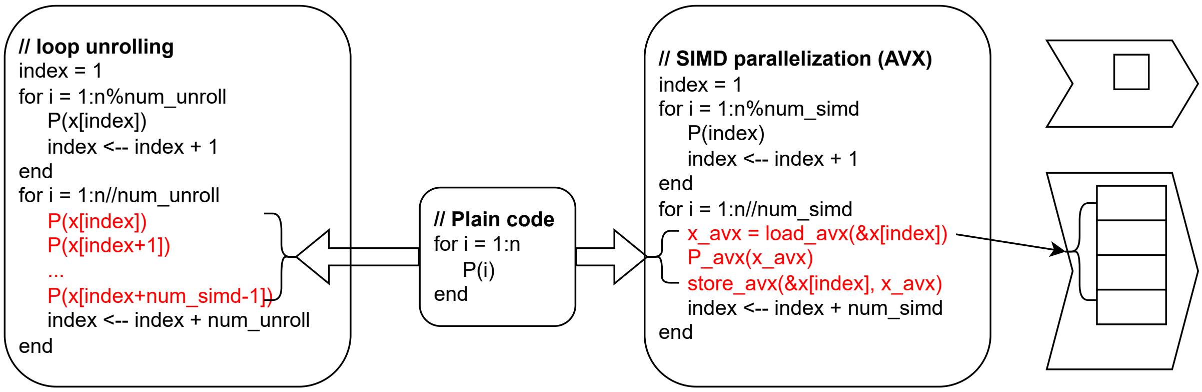


Figure 2. Demonstration for loop unrolling and SIMD parallelization. P means the operation for each time step (including all parameters and coefficients from some data), and x is the array-type data used for and updated by the operation. For the operation using SIMD parallelization, several entries of the data need to be loaded to some consecutive memory addresses, then the single instruction will be applied to these loaded entries simultaneously, and finally, store the result back to the data. Here for AVX2, the number of these double-precision entries for each iteration (num_simd) is 4.

abbr. of platform	machine	operating system	supported instruction sets
MBA	MacBook Air 2020 with 1.1 GHz 4-core Intel i5	MacOS 12	AVX, AVX2
MBP	MacBook Pro 2021 with 10-core M1 Max	MacOS 12	N/A
PC_Linux	AMD Ryzen 7 5800X 8-core 4.7 GHz	Ubuntu 22.04 LTS	AVX, AVX2
PC_Win	AMD Ryzen 7 5800X 8-core 4.7 GHz	Windows 11	AVX, AVX2

Table 1. Systems and hardware for numerical experiments

Algorithm 3 A fast biharmonic solver (simply supported boundary condition)

Input: $\mathbf{Q} \in \mathbb{R}^{(N_y-1)(N_y-1)}$, $\mathbf{V}_{kk}$ $(k = 1, 2, \ldots, N_y - 1)$, $\tilde{b} \in \mathbb{R}^{(N_y-1)(N_x-1)}$
Output: $\tilde{x} \in \mathbb{R}^{(N_y-1)(N_x-1)}$

```
function BiharmonicSolverSS(Q, V_kk, b̃)
    Initialize an empty matrix ṽ ∈ R^{(N_y−1)(N_x−1)}
    Solve Lv = b and get v̂:
    for j = 1 to N_x − 1 do
        b̂_{j*} ← (Q b̃_{*j})^T    ▷ i.e., (b̄_{*j})^T, and here we use Q instead of Q* since Q = Q*
    end for
    for k = 1 to N_y − 1 do
        v̂_{*k} ← SimplifiedThomasAlgorithm(V_kk, b̂_{*k})    ▷ Solve Γ_k v̂_{*k} = b̂_{*k}
    end for
    Solve Lx = v and get x̃:
    for k = 1 to N_y − 1 do
        x̂_{*k} ← SimplifiedThomasAlgorithm(V_kk, v̂_{*k})    ▷ Solve Γ_k x̂_{*k} = v̂_{*k}
    end for
    for j = 1 to N_x − 1 do
        x̃_{*j} ← Q(x̂_{j*})^T    ▷ i.e., Q x̄_{*j}
    end for
    return x̃
end function
```

In C++ implementation of the biharmonic solver, all matrices and vectors are using double-precision array data type, and matrices are stored by flattening them by column.

4.2 Alternative solvers

In the field of computational mathematics and scientific computing, people usually use either sparse matrix decomposition or iterative methods to solve linear systems [4], and the latter always deal with large-scale linear systems for storage or memory concerns and their time costs are generally higher than decomposition methods. The direct FFT-based solver is also another method to solve Poisson's equation with the Laplacian operator [3]. Therefore, considering the scale of the problem that is focused on in this paper, I'll only run the numerical experiments on the simple FFT-based solver and several decomposition-based solvers for comparison. Here, we use LU- and Cholesky-decomposition-based solvers for sparse matrices from Matlab and Eigen [13], a well-known high-level C++ library for linear algebra, for these numerical experiments.

5. NUMERICAL RESULTS

In this paper, four platforms including three machines and three operating systems listed in Table 1 are used for numerical experiments. The version of Matlab we used is R2022a, and plain code means no optimization techniques are used. All numerical results are cumulative time costs in seconds for 44100 iterations. All code should be compiled with at least -O3/-Ofast and -mavx2 [7] -march=native flags.

5.1 Comparisons between the biharmonic solver and alternative solvers

Here we compare the performance between the biharmonic solver and alternative solvers on PC_Linux platform. The results is shown in Table 2, where Matlab_B means directly solving the system $Bx = b$ using Matlab's default solver $B\backslash b$, Matlab_L^2 means solving two Laplacians using Matlab's default solver $L\backslash L\backslash b$, and Matlab_*_X means solving the linear system regarding B or two Ls by Matlab's default solver using decomposition X (chosen from LU or Cholesky). Eigen's abbreviations are similar to the above Matlab's. All C++ implementations using Eigen are complied with -O3 and -mavx2 -march=native flags since Eigen3.3 we used here supports both -O3 and AVX2 optimization. And we use four sizes of matrices, $(N_x - 1) \times (N_y - 1) = 14 \times 14, 16 \times 20, 23 \times 17, 25 \times 25$ for comparisons.

	14×14	16×20	23×17	25×25
plain code	0.916	2.051	2.202	4.686
loop unrolling	0.887	1.933	1.987	4.592
AVX2	0.551	1.007	1.246	2.385
plain code (-O3)	0.087	0.207	0.181	0.310
loop unrolling (-O3)	0.075	0.220	0.135	0.244
AVX2 (-O3)	**0.049**	**0.111**	**0.101**	**0.236**
Eigen_FFT	0.842	1.049	1.498	2.071
Eigen_B_LU	0.319	0.631	0.812	1.554
Eigen_B_Chol	0.227	0.416	0.527	0.930
Eigen_L^2_LU	0.357	0.854	0.943	1.997
Eigen_L^2_Chol	0.278	0.639	0.727	1.269
Matlab_FFT	1.030	1.960	2.062	2.630
Matlab_B	0.596	0.993	2.113	6.560
Matlab_B_LU	0.378	0.497	0.635	1.024
Matlab_B_Chol	0.363	0.403	0.523	0.919
Matlab_L^2	0.635	1.249	2.395	6.834
Matlab_L^2_LU	0.576	0.657	0.854	1.264
Matlab_L^2_Chol	0.548	0.603	0.732	1.043

Table 2. Numerical results of comparisons between the biharmonic solver and alternative solvers on PC_Linux. The first three rows are compiled without -O3 flag. Bold results are the best results for each column. Unit: second.

5.2 Comparisons of different optimization techniques for the biharmonic solver

Here we compare the performance of different optimization techniques for the implementation of the biharmonic solver on those four platforms. For the sake of brevity, we only show the results of three different optimization techniques compiled with -O3 flag with a grid size of $(N_x - 1) \times (N_y - 1) = 23 \times 17$. The results are shown in Table. 3.

[7] For AVX2. Choose -mavx for AVX and -mavx512f for AVX512.

	MBA	MBP	PC_Linux	PC_Win
plain code (-O3)	0.826	0.367	0.181	0.202
loop unrolling (-O3)	0.595	**0.242**	0.135	0.147
AVX2 (-O3)	**0.340**	N/A	**0.101**	**0.118**

Table 3. Numerical results of the biharmonic solver on different platforms. $(N_x - 1) \times (N_y - 1) = 23 \times 17$. Bold results are the best results for each column. Unit: second.

6. CONCLUSIONS

In this paper, we describe an algorithm for solving the discrete biharmonic system for modeling nonlinear plate dynamics based on a series of linear transformations and Thomas algorithm for tridiagonal systems. This method is good for optimization techniques on CPUs like loop unrolling or low-level SIMD parallelization using AVX intrinsics. At the scale of fast musical instrument simulation, the numerical results show that the C++ implementation of this method has better performance than other generally used methods for solving such linear systems like FFT-based and decomposition-based solvers, and indicate fast musical instrument simulation with nonlinear plate dynamics, which is actually used for real-time gong-like musical instruments synthesis based on the von-Kármán plate equation [8].

Acknowledgments

The authors would like to express their gratitude to Prof. Stefan Bilbao for his valuable contributions to this paper including the closed-form formulas presented in Eq. (7) and (8) and his constructive suggestions for modifications.

7. APPENDIX: PROOF OF LEMMA 2.1

Proof. First, it's obvious that $E = \{\mathbf{V}_{kk} \mid i = 1, 2, ..., N_y - 1\}$ has $N_y - 1$ distinct elements.

Then we only need to show the following equations,

$$\mathbf{A}q_k = \left(2\cos\left(\frac{k\pi}{N_y}\right) - 4\right) q_k, \quad 1 \le i \le N_y - 1,$$

where $q_{kj} = \sin\left(\frac{kj\pi}{N_y}\right), 1 \le k, j \le N_y - 1$.

For $2 \le j \le N_y - 2$, we have

$$\begin{aligned}
& \left(2\cos\left(\frac{k\pi}{N_y}\right) - 4\right) q_{kj} \\
= \quad & \sin\left(\frac{kj\pi}{N_y}\right)\left(2\cos\left(\frac{k\pi}{N_y}\right) - 4\right) \\
= \quad & 2\sin\left(\frac{kj\pi}{N_y}\right)\cos\left(\frac{k\pi}{N_y}\right) - 4\sin\left(\frac{kj\pi}{N_y}\right) \\
= \quad & \sin\left(\frac{(kj+k)\pi}{N_y}\right) + \sin\left(\frac{(kj-k)\pi}{N_y}\right) - 4\sin\left(\frac{kj\pi}{N_y}\right) \\
= \quad & \sin\left(\frac{(k(j-1))\pi}{N_y}\right) - 4\sin\left(\frac{kj\pi}{N_y}\right) + \sin\left(\frac{k(j+1)\pi}{N_y}\right) \\
= \quad & \mathbf{A}_{j(j-1)}\sin\left(\frac{(k(j-1))\pi}{N_y}\right) + \mathbf{A}_{jj}\sin\left(\frac{kj\pi}{N_y}\right) \\
& + \mathbf{A}_{jj}\sin\left(\frac{k(j+1)\pi}{N_y}\right) \\
= \quad & A_{j*}q_k.
\end{aligned}$$

For $j = 1$ or $N_y - 1$, notice that $\sin\left(\frac{k(1-1)\pi}{N_y}\right) = 0$ and $\sin\left(\frac{k(N_y-1+1)\pi}{N_y}\right) = 0$, which means the equation

$$\left(2\cos\left(\frac{k\pi}{N_y}\right) - 4\right) q_{kj} = \mathbf{A}_{j*}q_k$$

still holds for $j = 1$ and $N_y - 1$. Therefore, we have

$$\mathbf{A}q_k = \left(2\cos\left(\frac{k\pi}{N_y}\right) - 4\right) q_k,$$

which means E is the set of all $\mathbf{A}$'s eigenvalues, and q_k is the eigenvector w.r.t. $\left(2\cos\left(\frac{k\pi}{N_y}\right) - 4\right)$.

Notice that

$$\begin{aligned}
||q_k||_2^2 &= \sum_{j=1}^{N_y-1} \sin\left(\frac{kj\pi}{N_y}\right)^2 \\
&= \sum_{j=1}^{N_y-1} \frac{1 - \cos\left(\frac{2kj\pi}{N_y}\right)}{2} \\
&= \frac{N_y - 1}{2} - \sum_{j=1}^{N_y-1} \frac{\cos\left(\frac{2kj\pi}{N_y}\right)}{2} \\
&= \frac{N_y - 1}{2} - \frac{1}{2}\sum_{j=1}^{N_y-1} \mathbf{Re}\left(\exp\left(\frac{2kj\pi i}{N_y}\right)\right) \\
&= \frac{N_y - 1}{2} - \frac{1}{2}\mathbf{Re}\left(\sum_{j=1}^{N_y-1} \exp\left(\frac{2kj\pi i}{N_y}\right)\right) \\
&= \frac{N_y - 1}{2} - \frac{1}{2}\mathbf{Re}\left(\frac{\exp\left(\frac{2k\pi i}{N_y}\right) - \exp\left(\frac{2kN_y\pi i}{N_y}\right)}{1 - \exp\left(\frac{2k\pi i}{N_y}\right)}\right) \\
&= \frac{N_y - 1}{2} - \frac{1}{2}\mathbf{Re}\left(\frac{\exp\left(\frac{2k\pi i}{N_y}\right) - \exp(2k\pi i)}{1 - \exp\left(\frac{2k\pi i}{N_y}\right)}\right) \\
&= \frac{N_y - 1}{2} - \frac{1}{2}\mathbf{Re}\left(\frac{\exp\left(\frac{2k\pi i}{N_y}\right) - 1}{1 - \exp\left(\frac{2k\pi i}{N_y}\right)}\right) \\
&= \frac{N_y - 1}{2} - \frac{1}{2}\mathbf{Re}(-1) \\
&= \frac{N_y}{2},
\end{aligned}$$

for every $1 \le k \le N_y - 1$, where $i = \sqrt{-1}$. Thus, we have $||q_k||_2 = \frac{N_y}{2}$, and

$$\left[\frac{q_1}{||q_1||_2}, \frac{q_2}{||q_2||_2}, \cdots, \frac{q_{N_y-1}}{||q_{N_y-1}||_2}\right]^T = \mathbf{Q},$$

which leads to the following decomposition

$$\mathbf{Q}\mathbf{V}\mathbf{Q}^* = \mathbf{A}.$$

$\square$

8. REFERENCES

[1] T. von Kármán, "Festigkeitsprobleme im maschinenbau," *Encyklopädie der Mathematischen Wissenschaften*, vol. 4, no. 4, pp. 311–385, 1910.

[2] S. Bilbao, *Numerical Sound Synthesis: Finite Difference Schemes and Simulation in Musical Acoustics*. Chichester, UK: John Wiley and Sons, 2009.

[3] C. Temperton, "Direct methods for the solution of the discrete poisson equation: some comparisons," *Journal of Computational Physics*, vol. 31, no. 1, pp. 1–20, 1979.

[4] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU press, 2013.

[5] Z. Wang, M. Puckette, T. Erbe, and S. Bilbao, "Real-time implementation of the kirchhoff plate equation using finite-difference time-domain methods on cpu," *2023 Sound and Music Computing Conference (SMC)*, 2023.

[6] B. L. Buzbee, G. H. Golub, and C. W. N. R. work(s):, "On Direct Methods for Solving Poisson's Equations," *SIAM Journal on Numerical Analysis*, vol. 7, no. 4, pp. 627–656, 1970. [Online]. Available: http://www.jstor.org/stable/2949380

[7] L. Thomas, "Elliptic problems in linear differential equations over a network: Watson scientific computing laboratory," *Columbia Univ., NY*, 1949.

[8] S. Bilbao, C. Webb, Z. Wang, and M. Ducceschi, "Real-time gong synthesis," *submitted to DAFx*, 2023.

[9] S. Bilbao, "A family of conservative finite difference schemes for the dynamical von karman plate equations," *Numerical Methods for Partial Differential Equations: An International Journal*, vol. 24, no. 1, pp. 193–216, 2008.

[10] S. Bilbao and M. Ducceschi, "Fast explicit algorithms for Hamiltonian numerical integration," in *Proc. Eur. Nonlinear Dynamics Conf.*, Lyon, France, July 2022.

[11] S. Bilbao, M. Ducceschi, and F. Zama, "Explicit exactly energy-conserving methods for Hamiltonian systems," *J. Comp. Phys.*, vol. 427, p. 111697, 2023.

[12] W. W. Hager, "Updating the inverse of a matrix," *SIAM review*, vol. 31, no. 2, pp. 221–239, 1989.

[13] G. Guennebaud, B. Jacob *et al.*, "Eigen v3," http://eigen.tuxfamily.org, 2010.

On the acoustics of the concert kantele

Henna Tahvanainen
Department of Industrial Engineering, University of Bologna
henna.tahvanainen@unibo.it

ABSTRACT

The concert kantele is a plucked zither-like instrument with 36-40 strings. This paper presents a preliminary vibro-acoustical analysis of the concert kantele in terms of body admittance, modal parameters, and string plucking measurements. A short summary of the construction process of the instrument is given along with some initial discussion on the main design parameters affecting the acoustics of the instrument that pave the way for future research.

1. INTRODUCTION

The kantele represents a family of finger-plucked string instruments played in Eastern regions of the Baltic Sea including Finland, Estonia, Latvia, Lithuania, Poland, and Russia. The acoustics of the small five- to fifteen-string kanteles have been studied previously [2–7]. However, their construction tends to differ from the more modern kanteles with 30-40 strings presented in Fig. 1.

In order to extend the repertoire of the kantele, the concert kantele was initially developed by Paul Salminen in Finland [8, 9] and Hannes Wallen in the US in the early 20th century [10]. The kantele is tuned diatonically, and the invention of both Salminen and Wallen was to add a lever system similar to that of the concert harp to raise and lower the pitch of the string by a semitone allowing for various scales to be played with the concert kantele. The concert kantele has since become a popular instrument that can be studied at the university level, and multiple improvements have been introduced by makers thereafter [11]. The instrument is characterized by a bright sound and a long sustain with some beating. It is not a very loud instrument, and lately, both magnetic and contact microphones have become a standard installation to allow for ensemble playing in electronically amplified conditions. At the moment, there are two luthiers, Koistinen [1] and Lovikka [2], in Finland that make this kind of concert kanteles. The instruments studied in this paper are all built by Lovikka.

Structurally, the concert kantele resembles the zither with the wooden body of a curved trapezoidal shape. The strings (36-40 in total) are plucked with fingers (not with nails).

[1] https://www.koistinenkantele.com/in-english
[2] https://lovikka.com/en/

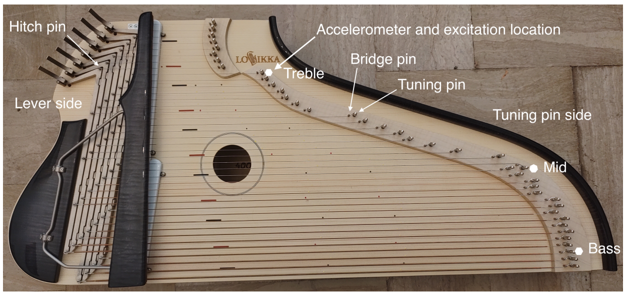


Figure 1: A modern concert kantele with the indicated measurement positions for the body admittance and the laser Doppler vibrometry measurements (white polygons).

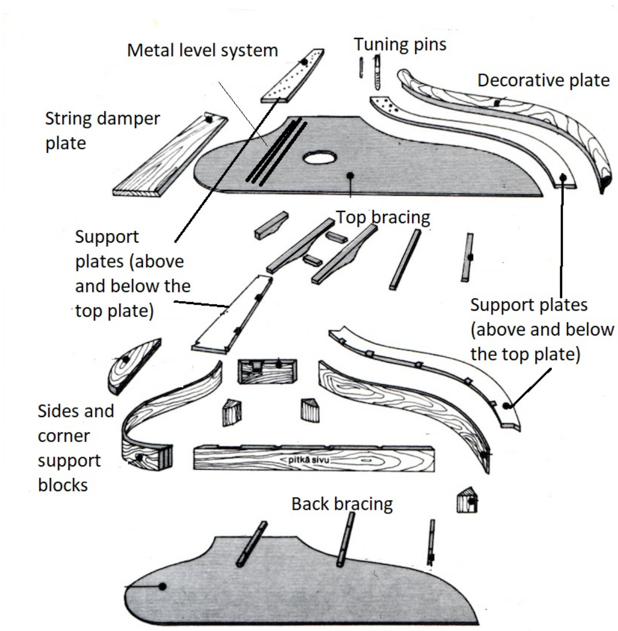


Figure 2: The basic construction of the concert kantele body with an exploded view, adapted from [1].

Figure 3: The different knottings used on the hitch pins of the concert kantele.

On one side, the strings are attached to individual tuning pins and under the tuning pins, bridge plates are glued on both above and below the top plate, as shown in Fig. 2. On the other side, the strings are attached to a mechanism of lever-actuated metal bars that elongate or shorten the string. There are seven bars, meaning that the same note in all octaves will undergo a simultaneous change of pitch (semitone up or down). Both single and wound strings are used to cover the notes from F1 (44 Hz) to C7 (2093 Hz). The total string tension can be as high as 8000 N.

Similar chromatic versions of the kantele are played across the Baltic states and Russia. But a crucial difference to these instruments is the lack of a distributed saddle as a coupling point between strings and body in the Finnish kantele, although a distributed bridge plate under the pins exists. The pitch change is also done differently: in Latvian and Lithuanian instruments, the lever mechanism is typically attached to each string individually similarly to the Middle Eastern kanoun, and in Estonia the chromatic strings are interplaced [12].

2. CONSTRUCTION OF THE CONCERT KANTELE

2.1 Body

The basic challenge of a concert kantele is to build a wooden body that can withstand the tension exerted by the strings and the weight of the metal lever system, and still vibrate satisfactorily. The top plate is only about 4-5 mm thick which means that the corners of the frame and the top plate under the tuning pins have to be reinforced.

The different parts of the concert kantele body are presented in Fig. 2. The building process for the body is described in Finnish by Alakuha [1]. In brief, the concert kanteles are made by first bending the edges to the right shape to form the frame and then glueing the braces and reinforcement plates onto the frame. There are two main reinforcement plates: one supporting the lever mechanism and one supporting the tuning pins. A special brace acting like a spring is attached between the two main plates to keep them apart under the tension of the strings (not drawn in Fig. 2). The top plate and back plate are then glued onto the frame. Then, the lever mechanism and the strings are installed. Fine-tuning the lever mechanism is one of the most time-consuming and crucial parts of the making of the instrument. For example, the use of the lever mechanism introduces changes in the tension and depending on the how the body is constructed, it yields differently to these changes. The error in tuning produced by the body yielding is somewhat compensated for in the design of the lever mechanism.

Luthier Jyrki Pölkki has introduced an alternative body construction where the top and the back plate are separated by an air gap [4]. It has been shown to produce an improved loudness [6] and low-frequency response [7]. The 40-string version of this instrument is currently available with individual tuning levers making the instrument essentially a concert kantele [3] .

[3] https://www.kurkikantele.com/kanteleet/

The top plate of the concert kantele is often made of Finnish spruce, but also rowan, ash, and maple are used. The material for the sides and the back plate is typically pine or alder, and a thin layer of flamed maple is used for finishing. Lately, thermally-aged wood has been used on most wooden parts to improve the ability of the instrument to withstand changes in environmental conditions.

As is typical for any wooden string instrument, the top plate and bracing material have an effect on the sound of the kantele. In general, the effects of specific material and design parameters of the kantele need more research, but based on the luthier's comments, some general observations can be made. Firstly, the material of the supporting plates appears to have an effect on the sound. Most commonly, European ash or maple is used, and the former yields a darker and warmer tone than the latter. Analogous to a guitar bridge (without the saddle), the supporting plate is the coupling point between the body and the string. As European ash tends to be a softer material than maple, the effect on the timbre could be explained by the internal losses of the wood. Furthermore, a thinner top plate and smaller sound hole radius have also been found to yield a warmer and darker tone, as expected from the acoustic principles of other string instrument designs that work as a coupled system of the top and back plate and the enclosed air with the sound hole. Similarly, the back plate thickness, material, and bracing affect the sound's darkness and warmness, which again can be understood based on research on other string instruments.

2.2 Strings

The kantele strings are made of steel and the thickness varies from 0.3 to 0.7 mm, and including winding for the lowest strings, the thickness is up 2.2 mm. The challenge is to have a near-uniform tension across to strings to have a similar feel for plucking, as well as, timbre. The concert kantele is typically tuned with A4=442 Hz.

On the tuning pin side, the strings pass around bridge pins before they are wrapped around the tuning pins. On the metal lever side, the strings are knotted around hitch pins. Several different knotting types can be used in a concert kantele, and examples from of a measured instrument are presented in Fig. 3. More knotting types are listed by Huotari [11].

The knotting gives rise to two different sounding lengths in the horizontal and vertical directions on the string and thus introduces beating in the sound [2]. The level of beating can be controlled with the knotting type, and the makers are using different knotting styles to market kanteles for different musical styles, such as classical or folk. The relationship between the knotting and the sound of the kantele merits further investigation in the future.

3. VIBRO-ACOUSTICAL MEASUREMENTS

Six concert kanteles from the same maker were measured with a normal playing setup on a table dedicated to the instrument in an anechoic room (4.3 m x 4.3 m x 4.5 m) that exhibits anechoic conditions down to 50 Hz. The signals

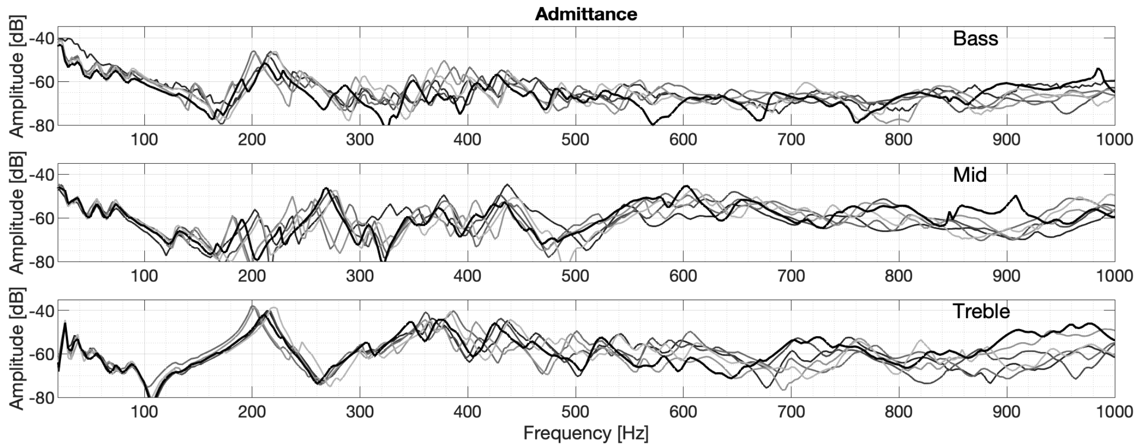


Figure 4: The concert kantele input body admittance of six instruments on bass, mid, and treble registers as indicated in Fig. 1. The instrument whose response is plotted with a black line is used for the subsequent modal analysis.

Mode	f [Hz]	ξ [%]	Mode	f [Hz]	ξ [%]
1	184.95	1.8212	21	568.79	1.2023
2	216.87	1.1805	22	590.12	1.0461
3	222.14	2.5364	23	639.26	1.2902
4	247.56	1.1031	24	647.68	1.9834
5	274.84	1.4537	25	684.73	1.6496
6	276.44	1.1001	26	703.27	2.6543
7	306.95	1.1792	27	738.39	2.6070
8	329.60	1.2348	28	764.35	0.7821
9	353.98	2.7303	29	795.12	1.7094
10	371.22	1.0189	30	837.07	1.1467
11	379.90	0.8711	31	847.55	1.7584
12	393.78	1.5775	32	899.19	1.0341
13	416.69	0.9431	33	904.23	1.9753
14	439.66	1.3951	34	943.45	1.0026
15	460.27	1.3363	35	983.40	2.5676
16	480.83	1.5325	36	993.15	2.0154
17	512.49	1.0399	37	1002.43	1.3948
18	513.56	0.9376	38	1009.77	1.0580
19	517.10	1.3494	39	1034.92	1.3455
20	541.54	1.7719	40	1043.29	1.4478

Table 1: The first 40 obtained eigenfrequencies f and their damping ratios ξ.

were recorded at 48 kHz with an RME UFX audio interface on a MacBook Pro. The kantele strings were damped with thin strips of felt, apart from the string that was being excited. A free-field microphone was set 1 m away from the instrument at an angle of about 45 degrees. The soundfield probe (Microflown 3D) was set on the sound hole of the instrument. The accelerometer (PCB 352C22SN) was set in the vicinity of the tuning pin corresponding to the string with the pitch A1, A3 and A5 to obtain the input admittances by hitting the corresponding tuning pins with an impulse hammer (PCB 086C01). The impulse hammer was found to provide a reasonable response up to 1 kHz. The accelerometer and the excitation locations are shown in Fig. 1. In addition, the aforementioned strings were plucked with the wire-breaking technique introduced by Woodhouse [13]. In this method, a thin copper wire (diameter 0.1 mm) is wrapped around the string and pulled until it breaks at about 1.3 N. The plucking was done perpendicularly to the soundboard at the plucking position of 2/5 of the sounding length of each string. The string lengths and thickness were in millimetres, respectively, A1: 736, 0.2+0.7 wound, A3: 671, 0.65, and A5:221, 0.4.

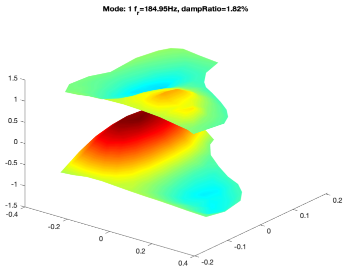
The mode shapes of one of the instruments were measured at the Institute of Nonlinear Mechanics at the University of Stuttgart. The eigenmodes and the eigenfrequencies of the concert kantele were measured with a laser Doppler vibrometer (LDV, Polytec PSV-500) and an automated impulse hammer. 401 measurement points were defined on a grid and three different driving points on the bridge plate corresponding to accelerometer locations in Fig. 1 were used since the wood was hard enough not to be dented by the impulse hammer. The kantele was resting on a metal grid resembling its playing table. The measurement equipment, setup, and analysis process are described in detail in [14] and more detailed analysis will follow in a subsequent manuscript.

4. RESULTS

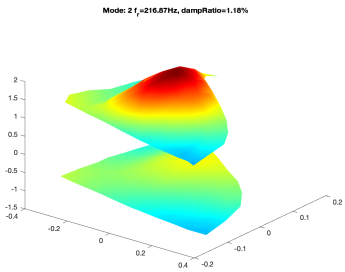
4.1 Body input admittance

The body input admittance for the six different kanteles is shown in Fig. 4, averaged over ten hammer hits on the bass, mid, and treble parts of the kantele, as shown in Fig. 1.

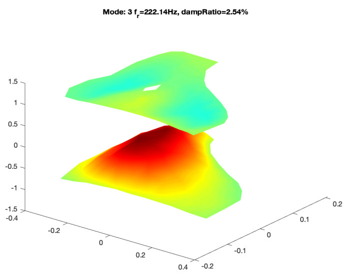
Apart from a few strong low-frequency resonances, the responses are fairly flat. It can be seen that the mid-part response is different at the low frequencies from the treble and bass parts. For these two responses, there seems to be a strong resonance between 200-220 Hz depending on the instrument, while for the mid-part the strongest resonance is at about 275 Hz. Similarly, for the bass and mid-part, smaller peaks at 180-190 Hz and 230-340 Hz can be observed. Another strong frequency range of vibration appears to be between 350-400 Hz.



(a) Mode 1: 184.95 Hz



(b) Mode 2: 216.87 Hz



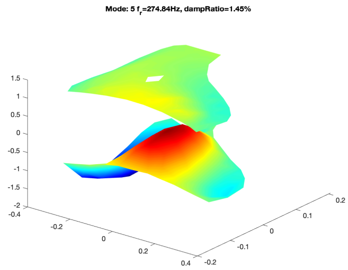
(c) Mode 3: 222.14 Hz

Figure 5: First three mode shapes of the concert kantele at about 200 Hz.

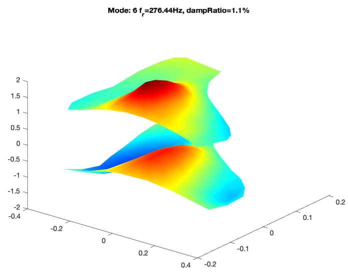
4.2 Modal parameters

The vibrational modes of the concert kantele plotted with the black line in Fig. 4 are listed in Tab.1 with corresponding frequencies and damping ratios. Roughly 36 modes can be found below 1 kHz, although at higher frequencies it becomes difficult to distinguish modes due to increasing damping ratios. The damping ratios are higher than those found in the acoustic guitar for example [14]. A possible cause for this could be the high string tension experienced by the soundboard.

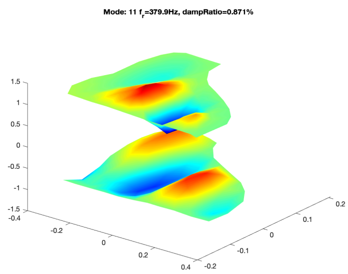
The visualisation of the first three mode shapes is presented in Fig. 5. They occur around 200 Hz, where the body input admittance shows peaks, especially in the bass and treble parts. The first mode (185 Hz) in Fig. 5a involves the top and back plate vibrating out of phase with one nodal line slightly diagonally across the body. On the second mode (217 Hz) Fig. 5b, mostly the top plate vibrates with one antinode, and the third mode (222 Hz) Fig. 5c, is the corresponding back plate mode with one antinode. With a crude theoretical estimate, the Helmholtz resonance of the instrument is at about 210 Hz, so it is likely that it couples with the top and back plates similarly to other string instruments with a sound hole.



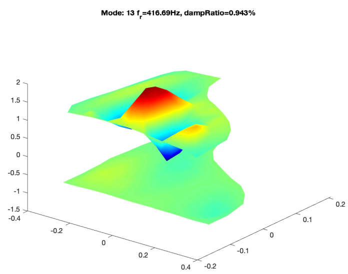
(a) Mode 5: 274.84 Hz



(b) Mode 6: 276.44 Hz



(c) Mode 11: 379.90 Hz



(d) Mode 13: 416.69 Hz

Figure 6: Other strong mode shapes.

Other strong mode shapes are shown in Fig. 6. The strong resonance in the mid-part body input admittance at about 275 Hz can be found in modes number 5 and 6 in Fig. 6a and 6b, respectively. Mode 5 seems to be back plate mode, and mode 6 is a combination of mode 5 and a top plate mode, where the maximum is located close to the long edge of the kantele, as opposed to mode 2, where the maximum is located towards the short side of the instrument. Also, the body input admittance indicates a peak at about 375 Hz. Fig.6c shows the corresponding mode number 11 around that frequency range. This can be identified as (3,0) mode of the top and back plate vibrating out of phase.

Generally, at higher order modes, the top plate seems to vibrate with smaller areas between the braces with various amplitudes, sometimes coupled with the back plate. The modes with the least amount of damping are associated with the movement of the tuning pin side of the bass register, as illustrated in Fig. 6d for mode number 13 identified as the top plate mode (4,0). Furthermore, the top plate around the metal lever system is not vibrating much, which yields asymmetric boundary conditions for the string attachment point on the left-hand and right-hand sides of the instrument.

4.3 String plucking

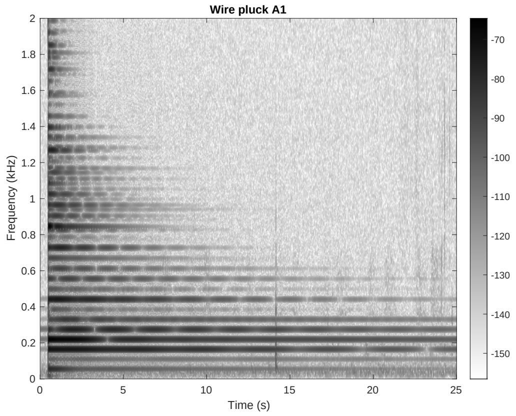
The timbre of the concert kantele is sensitive to the plucking point and plucking angle. There is also some variation in the decay times of the string overtones across different instrument specimens. However, some general observations present in all the measured kanteles can be made about the overtone series, and the spectrograms are shown in Fig. 7.

First of all, the decay times of the bass-register strings (over 30 s) are considerably longer than those of the mid- and treble register (about 8-25 s). Secondly, the partials are near-harmonic until the seventh partial, and inharmonic stretching is progressive at higher frequencies. For example, for the string A1, the inharmonicity reaches about 3% at 1 kHz. Thirdly, the plucking point at 2/5 of the sounding length would indicate that the fifth partial and its multiples should be missing as the plucking introduces an antinode at the nodal point of these modal shapes. However, this is not the case in any of the measured strings. Future research efforts should be targeted at looking into the possible nonlinearities of strings arising for example from the tension modulation driving force [3].

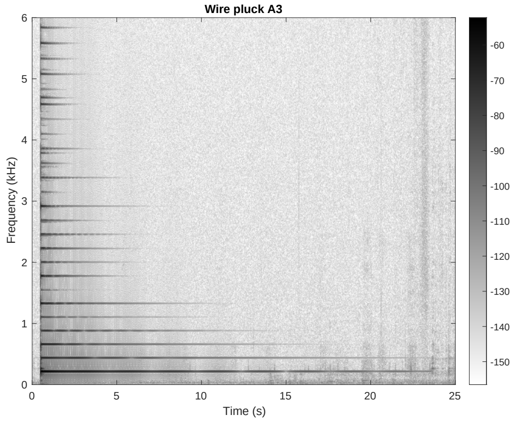
Figure 7a shows the spectrogram of the bass-register string (A1). This string is the richest in terms of the partials among the measured strings. It can be seen the first partial or the fundamental at about 55 Hz decays considerably faster than the rest. In some occasions, the noise levels at 50 Hz were interfering with the frequency bin of the fundamental, but in detailed inspection, they could be isolated. The second partial is weakly present, and the sound is dominated by the third, fourth and fifth partial. The frequency of the fourth partial fall in the range of the first top plate mode (mode 2), and indeed a strong attack sound can be seen in the spectrogram, as well. Some beating in the amplitude of some overtones is present.

Figure 7b shows an example in the mid-register, at about 220 Hz (A3). The sustain is not as long as in the case of the bass strings. The first partial is the strongest. Interestingly, the first partial is not affected by the top plate mode (mode 2) within the same frequency range, as the string attachment point is close to the nodal line of this mode shape. Instead, the body modes 1 and 5-6 are excited in the attack.

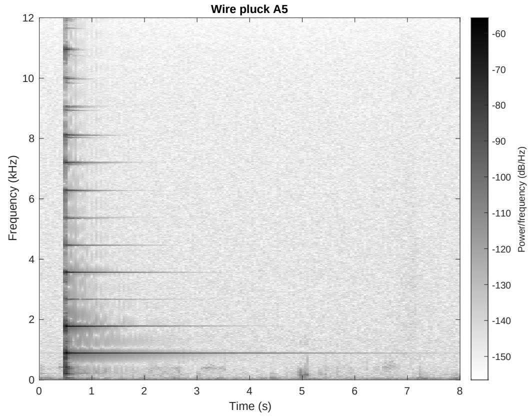
Figure 7c shows an example in the treble register, at about 880 Hz (A5). The first partial is the strongest and decays after about 8 s, which is considerably shorter than at the bass register. The second and fourth partial also stand out. The body modes 1 and 13 appear to be excited, well below the string fundamental frequency.



(a) String A1



(b) String A3



(c) String A5

Figure 7: Spectrograms of strings plucked with copper-wire.

5. DISCUSSION

It appears that the lowest modes of the concert kantele vibrate as a coupled system of the top plate, back plate and the enclosed air. The first top plate mode (mode 2) seems to be excited by the bass and treble parts of the tuning plate side. Its frequency varies between 200-220 Hz for different instrument specimens based on the body input admittance measurements. Other strong mode shapes involve the movement of the tuning pin side of the bass register.

The concert kantele string fundamental frequencies span eight octaves from 44 Hz to 2093 Hz. The first and strongest vibrational modes of the body are found between 180-400

Hz. This means that approximately the first two octaves of the instrument are not supported by any particular mode. This is illustrated by the lowest string that was plucked in this experiment, A1 at about 55 Hz, whose fundamental frequency decays very fast, and the third and the fourth partial at around 200 Hz dominate. The pitch perception of the lowest strings seems thus to rely on the phenomenon of the missing fundamental.

In the mid-and treble-register strings, the first string partial seems to be strongest. This is different from the five-string kantele where the second partial appears to be the strongest [2, 3]. The mid-and treble- register fundamentals seem to be well-supported by the flat response of the body. The treble-register strings are able to excite some of the lower body modes that fall below the frequency of the string fundamental. The short decay times at the treble-register emphasise the impression of an strong attack sound compared to the bass-register strings.

The sustain of the string vibration varies across octave registers from about 8 seconds to over 30 s in the measurements. Such variety means that the player must control the string vibrations differently in different registers. In fact, learning the use of different damping mechanisms (finger, arm, damping board) constitutes an essential part of the playing technique of the instrument.

Finally, the string overtone series is slightly inharmonic as expected. The strings do not seem to follow the principles of the linear string regarding the plucking position and missing overtones. The beating caused by the knotting is occasionally present, but both beating and the content of the overtone series seem to depend on the plucking angle, which was not studied here.

Multiple aspects of the concert kantele should be studied further. These include a larger set of string measurements with varied plucking angles, modal analysis of several specimens and the influence of several design parameters, as well as, the effects of the lever system of the body yield and vibration.

6. CONCLUSIONS

The concert kantele body exhibits a few strong resonances in the low frequencies. Otherwise, the body input admittances are fairly flat and the eigenmodes have fairly high damping ratios, probably to due to the high string tension and the sturdy construction of the body needed to support it. The first set of modes is particularly interesting as it involves both the top and back plate vibrating with one nodal line usually associated with higher mode numbers. Generally, the mode shapes are split into many antinodal areas on the top and back plates, and the lowest damping values are associated with antinodes at the bass register on the tuning pin side. The metal bar side barely vibrates, which creates asymmetric boundary conditions for the string termination points.

The different string registers vary considerably in decay times. The overtones series is slightly inharmonic. The fundamental frequency seems to dominate the string vibrations apart from the bass register where the fundamental dies quickly, and the second harmonic is not very powerful either. Thus, at these bass-register strings the pitch perceptual is likely to rely on the missing fundamental phenomenon. Because of the complex nodal lines of the body modes, the excited body modes by the string vibrations vary according to the string attachment point in the body.

Acknowledgments

This work has received funding from the Finnish Furniture Fund and the Finnish Cultural Foundation. The author would like to thank master luthier Pekka Lovikka for crafting the instruments and discussion, and MSc. Alexander Brauchler and DSc. Pascal Ziegler for hosting the author for the LDV measurements and providing the analysis, and prof. Ville Pulkki for updating the measurement equipment at Aalto University.

7. REFERENCES

[1] J. Alakuha, *Kanteleen rakennusopas*. Kansanmusiikki-instituutti, 1981.

[2] M. Karjalainen, J. Backman, and J. Polkki, “Analysis, modeling, and real-time sound synthesis of the kantele, a traditional finnish string instrument,” in *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1. IEEE, 1993, pp. 229–232.

[3] C. Erkut, M. Karjalainen, P. Huang, and V. Välimäki, “Acoustical analysis and model-based sound synthesis of the kantele,” *The Journal of the Acoustical Society of America*, vol. 112, no. 4, pp. 1681–1691, 2002.

[4] J. Pölkki, C. Erkut, H. Penttinen, M. Karjalainen, and V. Välimäki, “New designs for the kantele with improved sound radiation,” in *Proceedings of the Stockholm Music Acoustics Conference*, 2003, pp. 133–136.

[5] A. Peekna and T. Rossing, “The acoustics of carved baltic psaltery,” *The Journal of the Acoustical Society of America*, vol. 115, no. 5, pp. 2565–2566, 2004.

[6] H. Penttinen, C. Erkut, J. Pölkki, V. Välimäki, and M. Karjalainen, “Design and analysis of a modified kantele with increased loudness,” *Acta Acustica united with Acustica*, vol. 91, no. 2, pp. 261–268, 2005.

[7] H. Tahvanainen, J. Pölkki, H. Penttinen, and V. Välimäki, “Finite element model of a kantele with improved sound radiation,” in *Proceedings of the Stockholm Music Acoustic Conference, Stockholm, Sweden*, 2013, pp. 193–198.

[8] P. B. Salminen, “Accessory for kantele and similar instruments,” Patent, Patent FI8748A, 1921.

[9] ——, “Tuner for kantele and similar instruments,” Patent, Patent FI11483A, 1925.

[10] H. Wallen, “Harp,” Patent US2 169 938A, Patent US2169938A, 1937.

[11] S. Huotari, *Kanteleen huoltokirja*. Kanteleliitto, 1985.

[12] A. Peekna and T. D. Rossing, "Psalteries and zithers," *The Science of String Instruments*, pp. 99–122, 2010.

[13] J. Woodhouse, "Plucked guitar transients: Comparison of measurements and synthesis," *Acta Acustica united with Acustica*, vol. 90, no. 5, pp. 945–965, 2004.

[14] A. Brauchler, P. Ziegler, and P. Eberhard, "An entirely reverse-engineered finite element model of a classical guitar in comparison with experimental data," *The Journal of the Acoustical Society of America*, vol. 149, no. 6, pp. 4450–4462, 2021.

IS THE WOODBLOCK AN AEROPHONE?
A discussion among acoustics and organology

Leonardo Fuks
UFRJ-School of Music
fuks.leonardo@gmail.com

ABSTRACT

The woodblock is one of the simplest professional percussion instruments and is broadly regarded as an idiophone. It is usually made from a single block of wood, and widely employed in concert music, pop bands, percussion ensembles, Latin music, among others. Most references designate it as an unpitched instrument, despite its rather clear tonal sound, particularly if played in contrast with one or more woodblocks. Depending on the maker and model, it is finely adjusted to match a given note. Preliminary analysis and measurements indicate that the woodblock's nominal pitch is directly related to the acoustical resonance of the instrument cavity. We applied simplified models of Helmholtz resonator and open-closed pipe for estimating the resonances and nominal pitch of a woodblock and compared them to the experimental data obtained from a commercial instrument and a handmade one with close measurements. We found two main lower resonances, one related to the air cavity and another to the top plate. Both of them seem to be relevant in the sound produced. We might describe the woodblock as containing a vibrating top plate and an air chamber with a wooden flexible layer that is excited by the striking of a mallet, thus, placing it in the hybrid category of idiophones-aerophones. A new technique for the playing of the woodblock is proposed, based on the paper results.

INTRODUCTION

The woodblock is one of the simplest professional percussion instruments [1]. It usually consists of a solid block of wood, with a large longitudinal slit that creates a thin surface on the top and an air cushion, which is regarded as a resonant cavity, see Figure 1 [2]. It is compared to African slit drums in ethnomusicological and organological studies [3,4,5]. It is broadly regarded as an idiophone, requiring a comparatively simple performance technique, being widely employed in concert music, pop bands, percussion ensembles, and Latin American music, among others. Apart from being a musically relevant and commercial instrument, it has not received much attention in the realm of acoustics.

The name “woodblock” conveys an image of a sturdy and bulky instrument that vibrates and irradiates sound when struck by a mallet.

The one employed in this study, see Figure 1, is a Ron Vaughn Signature #7 Genuine Mahogany Woodblock, pitch Db5. It costs about U$ 250.00 and weighs ca. 2000 grams. The instrument was preliminarily tested by the author, who compared its sound to those of a piano. It was noticed that it simultaneously contained the pitches of Db_5 (as informed by the maker), but also an E_5, resulting in a minor third. This was informally shown to other musicians, who also heard this dual pitch of the instrument.

Figure 1. The Ron Vaughn #7 Woodblock in Mahogany, with nominal pitch Db5, Block Dimensions: 10 cm x 14 cm x 33 cm. Slit size, approx. 2 cm x 12.5 cm x 29 cm, with internal volume of ca. 700 ml. Weights ca. 2 kg

This “double resonance” raises a question of its nature. This paper discusses if a thin slice of air, represented by the block slit, with a length of 29 cm, a width of 2 cm and a depth of 12.5 cm, weighing less than 1.0 gram, might be the instrument’s primary sound oscillator, thus making it an aerophone.

1.1 *Similar Instruments:: tuning fork resonant box and boomwhackers*

For this discussion, we have selected two instruments that may have some common points with the woodblock, in terms of sharing characteristics of the idiophone and the aerophone. Rudolf Koenig, trained as a fine violin maker, was the main designer and manufacturer of tuning forks, in the second half of the 19th century. He supplied Hermann Helmholtz with acoustical instruments and was probably the one who developed the wooden resonant box for the tuning fork, see Figure 2. These boxes were made of finely-grained spruce with a light varnish and mahogany veneer on the side [14].

Figure 2. A tuning fork mounted on a wooden resonant box. The largest boxes are closed on one side and open on the other side. They behave as resonant pipes, with a wavelength of about 4 times the box's internal length. The boxes are finely adjusted to the tuning fork frequency.

If the resonant box surface is struck by a soft mallet, it will produce a clear tonal sound, coinciding with the tuning fork pitch. In this case, if regarded as a musical instrument, the resonant box should be primarily considered an aerophone. But, due to the "wooden" sound and making, as well as the percussive performance method, it might also tend to be defined as an idiophone.

Another emblematic case is the Boomwhacker, a very successful and commercial instrument, frequently employed in musical education. It consists of a flexible plastic pipe that is stiff enough to establish standing waves within, and soft enough to be hit against each other, or even against the player's bodies, with no risk of harm, see Figure 3.

Figure 3. A set of Boomwackers: flexible cylindrical pipes that are excited by percussive gestures, mainly by concussion, by hitting on the body, or by stamping on the ground. The pitched sounds derive from the acoustical lengths of the air columns, thus, characterizing an aerophone.

The use of a plastic pipe for the production of percussive sounds is widely disseminated, particularly in popular music, educational instruments and experimental music. Ensembles like Uakti and Les Luthiers have been using them since the 70's, at least. Blue Man Group employs percussive pipes since their foundation in 1987. However, it is hard to find references for that in papers in organology or acoustics. The most common name for such an instrument that we found was "thongophone"; it is not even an appropriate term, because it refers to the object used as a "mallet", a thong - a type of rubber footwear, also known as flippers or flip flop. Similarly to the udu, Figure 4, the thongophones and Boomwhackers are mainly aerophones but may be also included in the idiophone category. One could argue that, if the thongophone is excited by the lateral striking against its walls, the sound will become less clear than that of an aerophone, but still the air column remains a primary oscillator.

1. The Sound of the Woodblock

Although the woodblock is not often referred to by the musicological literature, an interesting study by Reymore and Huron [6] on musical instrument timbre mentions it with an emphasis on several perceptual dimensions, using 23 experienced musicians. The experimental paper compared 20 different instruments, including 7 percussions (woodblock, bass drum, snare drum, cymbals, timpani, triangle, vibraphone, piccolo, flute, oboe, clarinet, alto saxophone, English horn, French horn, tuba, bagpipes, kazoo, banjo, piano, harp). The method consisted of performing interviews with the musicians, asking them to rate the instruments according to 77 different perceptual categories.

The results, obtained from a huge amount of data generated, indicated that the woodblock was, among all instruments, the one that received indications of being the instrument with the **highest** rate of association to the terms: focused/compact, hollow, simple, and woody.

Also, the woodblock was the instrument with the **lowest** rate of association with the following terms: dramatic/expressive, singing/voice-like, metallic, mournful/wailing, resonant/vibrant, rich/complex, ringing/long decay, sad/melancholy, sustained/even, watery/fluid, wavy/undulating.

The investigation above clearly indicates the very special way the woodblock is received by the expert listener.

2. Organological aspects

The pioneer Belgian organologist and acoustician Victor Mahillon (1841-1924) divided the mechanical instruments into four categories [7], particularly: (1) the autophones ('self-sounding instruments'), "whose material is sufficiently rigid and at the same time sufficiently elastic to undergo periodic vibration"; and (4) wind instruments ('instruments a vent') "in which a column of air vibrates" [8]. It would be very reasonable to include in group 4 the instruments with a vibrating air cavity, such as the iconic Helmholtz resonator and the popular ocarina.

According to the systematic musicologists from Berlin, Curt Sachs (1881-1959) and Erich von Hornbostel (1877-1935), who expanded Mahillon's work, the categories above were renamed (1) idiophones [8,9], and (4) aerophones. Idiophones are "instruments made of naturally sonorous materials not needing any additional tension as do strings or drumskins". They are self-resonant, producing sounds when struck or otherwise excited mechanically. The group includes virtually all percussion instruments that are not of the membranophone family. (4) Aerophones include what are usually called 'wind instruments,' with the addition of a few instruments with a different acoustical principle called 'free aerophones'. Aerophones usually employ an air column (or air cavity, we might add), driven by a suitable source, to produce and sustain sound [10].

Summarizing, what defines an organological category is the primary sound generator. In idiophones, the vibrating body of the instrument directly irradiates acoustical energy to the environment. In aerophones, the air column, or air cavity that is somehow excited, transfers sound oscillations to the medium outside the instrument's borders. It is also an important issue the way instruments are excited by the player. Therefore, within idiophones, they are divided into: idiophones struck together (concussion instruments); struck idiophones, when hit by a mallet or by the hands; stamped idiophones (hit on the ground in vertical motion), including the large slit-drums; shaken idiophones; scraped idiophones, and, plucked idiophones. The woodblock would be traditionally included in the struck idiophone group, and it would hardly be excited by another method. About the excitation tool, larger woodblocks will require softer mallets than smaller ones. However, little is found in terms of organological papers about the woodblock.

2.1. Instruments from more than one category

Some instruments may easily belong to two organological categories. For instance, the Brazilian tambourine, the "pandeiro", is an idiophone and a membranophone, since it contains several pairs of small cymbals, and a leather or plastic drumhead, being normally activated jointly during the performance. This is due to its proper design. In other cases, drums are alternatively played by striking on the hoop, i.e., the metal or wooden frame that holds the membrane. The same occurs with a guitar when it is tapped on the soundboard, a very common technique in several genres, from flamenco to contemporary classical music. These performance techniques obviously bring idiophonic features to the instruments, but it might sound awkward, especially for their players, if a classical guitar or a snare drum were called idiophones for that reason. On the other hand, a Jew's harp, traditionally regarded as an idiophone, should be included in the aerophone category [11]: it would be enough to observe that a very common and traditional playing technique of the Jew's harp entails the vigorous blowing and drawing of air through the vibrating tongue, generating a type of sound that coincides with that of a free reed, just like a mouth harmonica. The fact of the Jew's harp sound resonates in the mouth cavity, per se, doesn't place it in the aerophone category, because the vocal tract might be considered a secondary and not a primary resonator, in some cases. Finally, the Nigerian **udu**, a large hollow vessel with a neck and a hole, equivalent to a Helmholtz resonator, widely used by percussionists around the world, is both regarded as an idiophone and an aerophone [12], see Figure 4. The players explore the sound possibilities of the instrument's body in ceramics (idiophone) and the various effects of the air cavity (aerophone).

Figure 4. Udu made by Latin Percussion Company, inspired by the Nigerian traditional instrument. It is both an idiophone and an aerophone.

Another "instrument" that combines idiophone and aerophone features are the clapping hands [14]. The sounds created are the result of an explosion of air, that excites the resonances of the cavity formed by the fingers and hand palms.

A poor assumption, however, would be that of associating the xylophone, the marimba, or the vibraphone with this hybrid category of idiophone-aerophone. It is clear that the pipes present in those instruments are acoustically connected to the vibrating bars, and serve as a secondary, but not as a primary resonator.

2.2. **Woodblock Specifications**

Woodblocks are normally described by their sizes and pitches. Table 1 below gives information about some of Ron Vaughn's woodblocks, whose collection ranges over 17 pitches from D$\#_6$, to B_4.

Model	Wood species & size	Pitch	Best Mallet
W-0.8	N. Amer. Black Cherry, 1 5/8" x 1 3/4" x 4 3/4"	D# 7	PBM-1R
W-0.9	N. Amer. Black Cherry, 1 5/8" x 1 3/4" x 4 3/4"	D 7	PBM-1R
W-1	N. Amer. Black Cherry, 1 5/8" x 1 3/4" x 4 3/4"	C# 7	PBM-1R
W-1.2	N. Amer. Black Cherry, 1 5/8" x 1 3/4" x 4 3/4"	B 6	PBM-1R
W-1.3	N. Amer. Black Cherry, 1 3/4 x 2" x 5 1/4"	A 6	PBM-1R
W-1.4	N. Amer. Black Cherry, 2 1/8" x 2 1/4" x 6"	G 6	PBM-1R
W-1.5	N. Amer. Black Cherry, 2 1/8" x 2 1/4" x 6"	F 6	PBM-1.5R
W-2	N. Amer. Black Walnut, 2 3/8" x 2 3/8" x 7"	E 6	SBM-2R
W-3	N.Amer. Hard (rock) Maple, 2 3/4" x 2 3/4" x 9"	C# 6	MBM-2R
W-4	Genuine African Mahogany, 3" x 3 1/4" x 9 1/4"	B 5	MBM-2R
W-4.5	Genuine African Mahogany, 3" x 3 1/2" x 3 3/4" x 10"	A 5	LBM-3R
W-4.7	Genuine African Mahogany, 3 3/4" x 3 7/8" x 10 1/2"	G 5	LBM-3R
W-5	Genuine African Mahogany, 3 7/8" x 4 5/8" x 10 5/8"	F 5	LBM-3R
W-6	Genuine African Mahogany, 3 7/8" x 5 1/4" x 12 3/8"	Eb 5	LBM-3R
W-7	Genuine African Mahogany, 3 7/8" x 5 5/8" x 13"	Db 5	LBM-3R
W-8	Genuine African Mahogany, 3 7/8" x 5 5/8" x 13 1/4"	B 4	LBM-3R

Table 1. Data provided by maker Ron Vaughn, relating models, dimensions, corresponding pitches, and recommended mallets.

3. Acoustical Aspects

4.1. **Helmholtz resonance**

The Helmholtz resonator represented by the slit has no constructed neck to delineate the vibrating mass; still, the air layer at the outer face of the slit may be regarded as a "virtual" neck, initially estimated as 0,85 times the radius of the circular hole that matches the area of the slit opening [13]. This would be a rough approximation, taken from a cylindrical neck configuration, and further modeling would be needed for the slit geometry in the future. We will calculate the value and compare it with the experimental values to check for coherence.

$$f_h = \frac{c}{2\pi}\sqrt{\frac{1.85\,r}{V}} \qquad (1)$$

f_h is the lowest estimated resonance of the woodblock slit; r is the radius in meters of the circular hole that matches the area of the slit opening; V is the volume of air inside the slit, in cubic meters.

For the slit frequency calculation, the value in meters of r in formula (1) should be replaced as follows, S in square meters:

$$r = \sqrt{\frac{S}{\pi}} \qquad (2)$$

In the case of the present woodblock, given the dimensions L for the slit length, W for the slit width and D for the slit depth, the volume V is the product of the three dimensions, while S is the product of W and L.

4.2. **Pipe Resonances**

The woodblock slit, with a rectangular cross-sectional geometry of five walls, might, in principle, accommodate two types of pipe resonances. One would be the open-closed pipe from the external part to the slit depth. The other, the parallel narrow faces with a distance of the slit length.

The open-closed mode would correspond to one-fourth of the wavelength. Therefore, this mode, if it significantly occurs, corresponds to a wavelength that measures four times the slit depth.

In the same way, the hypothetical longitudinal mode will have a wavelength corresponding to two times the slit length.

4.3. **Wood Plate Resonances**

The woodblock may be schematically separated into two contrasting wooden layers: the top plate, with a thin thickness (ca. 15 mm in our #7 model), and the base plate, with a thickness of about 60 mm. It is reasonable to hypothesize that the base functions as an inert and stiff mass of wood, to ensure firmness to the top plate oscillations. It is very likely that the reduction or increase of the base thickness by 20%, for instance, will have little to no impact on the woodblock's sound. The top plate, with an approximately rectangular shape, being clamped at three sides to the block, having just one "free" side, that of the slit opening, represents a rather complex configuration, not available in most acoustics and vibration texts.

Therefore, the material mechanical properties, the actual design of the plate, and its connection to the remaining block are a challenge for the modeling of the top plate's vibratory behavior with the excitations produced by the mallet action. Anyway, we could suppose that the top plate will have a very rich response in frequency and that these vibrations will be somehow coupled to the air slit vibrations, probably showing relevant co-oscillations. These complex phenomena are beyond the scope of this paper. A preliminary simulation has been done, using a finite elements method software, applying the dimensions and material of the woodblock, shown in Figure 12. Further work is required to adjust the simulation to the experimental data.

4. Exploring the Handmade Woodblock

A woodblock similar to the Ron Vaughn #7 was built by the author with plywood, see Figure 5.

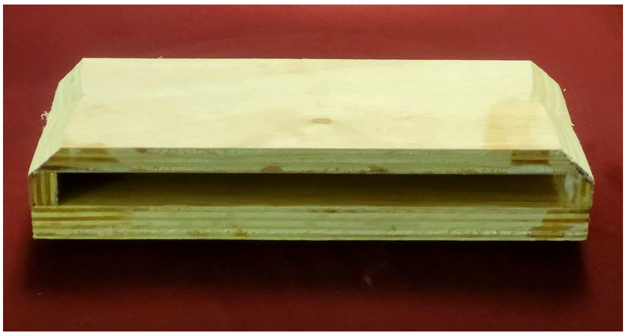


Figure 5. Woodblock in plywood, here called Plywoodblock. The top plate is 15 mm thick and the base is 50mm (10 mm thinner than the original 60 mm). The overall dimensions and the slit size are almost identical to the one from Ron Vaughn #7. The slit volume in both woodblocks is ca. 700 ml.

After gluing the pieces, the sound was muffled, probably due to internal gaps. About 50 grams of epoxy were used to seal the slit, which improved the sound and made the main resonant frequency rise by a considerable interval. It must be noted that the Plywoodblock does not sound as good as the #7 woodblock. If the base plate was thicker, it might yield a more "woody", "focused" and "hollow" sound. Probably, woodblock makers optimized the dimensions by progressively reducing the thickness for a given pitch. This is a very simple task and it saves wood and shipping costs.

The most affirmative fact is that both woodblocks responded with the same resonant frequency, when excited by a "damping strike", i.e., with the cotton thread mallet head being directly held with the hand and squeezed against the top plate after striking it. This type of excitation kills most of the top plate's vibrations, enhancing the slit resonance.

5. Measurements and Results

The experimental procedures consisted of recording the sounds from three types of excitation of both woodblocks - the #7 Ron Vaughn and the handmade Plywoodblock- (1) playing the woodblocks with a medium-soft mallet (similar to a LBM-3R by Ron Vaughn); (2) blowing air like "playing a pan flute" in the middle region of the woodblocks, with a flow perpendicular to the slit walls; (3) playing the woodblocks with a "damping strike", as described above, keeping the mallet in contact after the shock. A ZOOM H4 digital microphone was directly connected to the computer. The signals were processed by the Praat (praat.org) software package, to obtain the main spectral peaks.
Initially, the slit's resonance frequency was calculated by formulas 1 and 2, using the dimensional data. The estimated Helmholtz frequency in both cases was 585 Hz. The estimated natural frequency of the slit as an open-closed pipe corresponds to approximately 690 Hz, without any end correction. The hypothetical longitudinal resonance of the slit, corresponds to ca. 595 Hz. We observe that the three resonances are situated in the interval between 585 Hz and 690 Hz.
The spectra, as expected, are very rich in frequencies over a wide range. The most reliable natural frequencies of the woodblocks are the ones obtained by the "damped strike", represented as "slit responses" on Figures 10 and 11. The #7 woodblock presents a 654 (E_5 +14 cents) Hz peak, while the Plywoodblock peaks at 637 H (Eb_5 -40 cents).
The air flow noise response (Figures 8 and 9) reveals resonances pairs of 507 Hz (B_4 -45 cents) and 650 Hz (E_5 +24 cents) for the #7 woodblock; and 537 Hz (C_5 -45 cents) and 640 Hz (Eb_5 -49 cents) for the Plywoodblock.
The actual playing of the woodblocks results in 560 Hz (Db5 -17 cents) and 637 Hz (Eb5 -41 cents) for the #7 woodblock and results in 507 Hz (B4 -45 cents) and 650 Hz (E5 +24 cents) for the Plywoodblock. As expected, the main frequency of the #7 woodblock corresponds to the Db 5. The other pitch initially perceived probably refers to the Eb, but this deserves further investigation, since we noticed a minor third interval, not a major second on the #7 woodblock.

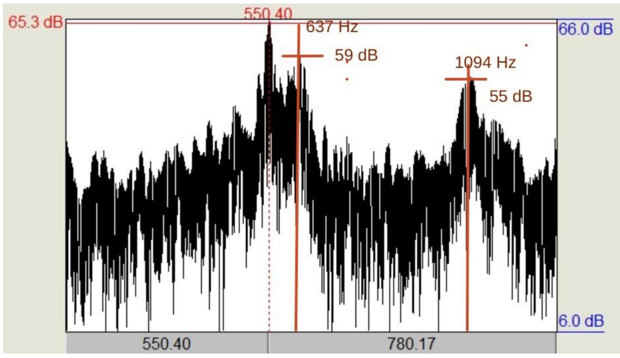


Figure 6 . Playing the #7 Woodblock. The spectrum shows resonant peaks at 550 Hz, 637 Hz and 1094 Hz.

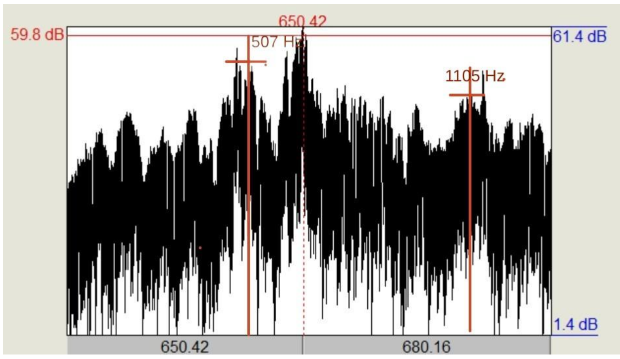


Figure 7. Playing the Plywoodblock - The spectrum shows resonant peaks at 507 Hz, 650 Hz and 1105 Hz. These values are close to the ones obtained with the #7 woodblock. They have very almost identical dimensions at the slit.

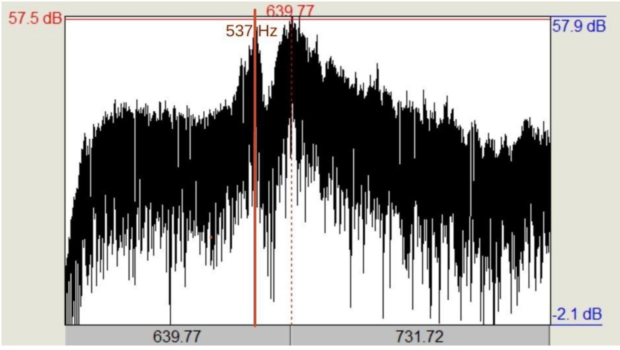


Figure 8. Blowing into the #7 Woodblock - The spectrum shows resonant peaks at 537 Hz and 640 Hz. The values are very close to the resonance peaks for the percussive response, see Figure 6.

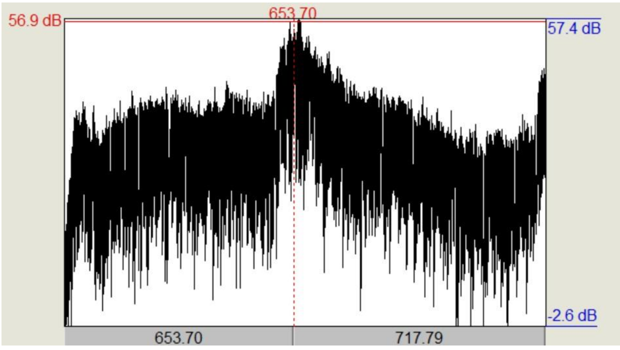


Figure 9. Blowing into the Plywoodblock- The spectrum shows one resonant peak at 654 Hz. The value is very close to the second resonance peak for the percussive response, see Figure 7.

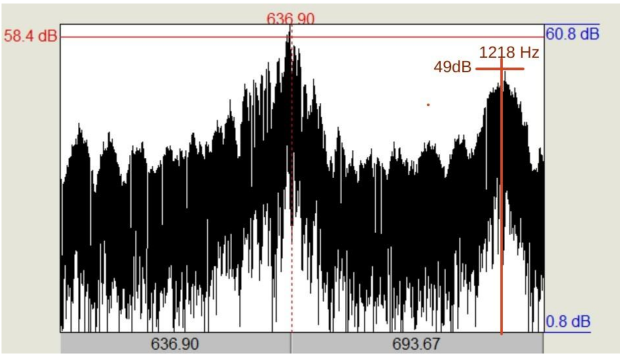


Figure 10. Slit Response - #7 Woodblock - The spectrum shows one resonant peak at 637 Hz and another at 1218 Hz. This type of excitation, with the damping of the top plate, should reveal only the air resonances.

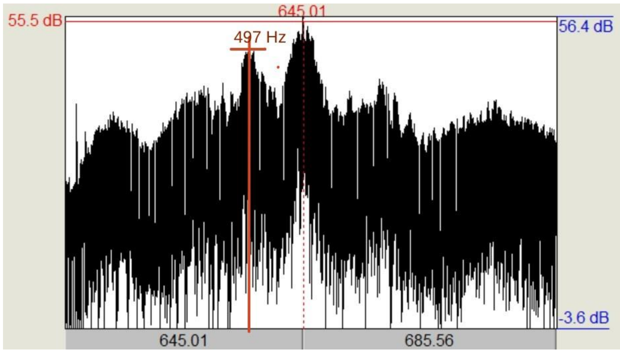


Figure 11. Slit Response - Plywoodblock - The spectrum shows one resonant peak at 497 Hz and another at 645 Hz.

6. Discussion

6.1. Pitched or unpitched instrument?

This criterion for the distinction among instruments is frequently employed by musicians and musicologists [13] but it is not technically simple to define. Many references designate the woodblock as an unpitched instrument [13, 15,10], in spite of being reported as carrying a clear tonal sound by others [2,16].The attribution of a pitch to a sound, if it is complex and made up of non-harmonic components, may strongly depend on context and task [17,18]. Therefore, if one hears two or more woodblocks, for instance, the perception of pitch may be readily available, because the distinctive tone will be considerably more salient. The fact of this particular maker, Ron Vaughn, trading pitched woodblocks for more than 50 years, and many composers writing for them, must be obviously taken into account. Therefore, it is reasonable to admit that the woodblock is a pitched instrument, regardless of the possible contradictions in this terminology.

6.2. The interaction between top plate and air cavity

Our exploratory study was confined to the slit resonances of a single woodblock. The measurements performed strongly indicate the natural frequencies of the slit, in the region of 585 Hz to 690 Hz to determine the overall sounds produced and perceived by the woodblocks. These procedures should be applied to the whole set of woodblocks available. This will show a full panorama of the instrument. It is also plausible that the top plate interacts with the slit vibrations, producing co-oscillation and, probably some interesting effects of amplitude modulation, with possible impacts on the perceptions of those sounds. The probable occurrence of amplitude modulation will benefit from recordings done in an anechoic chamber. Also, the sound signals may be manipulated to have their amplitude envelopes, which may provide useful insights for the definition of pitch identification in percussion instruments.

The top plate must be further investigated with finite elements methods in the future. Some preliminary simulations done with the software SimSolid indicate that the first mode may have a frequency close to the first resonance peak, see Figure 12. The possible matching of the air cavity and the top plate resonances seems to be a very convenient improvement to the instrument's efficiency.

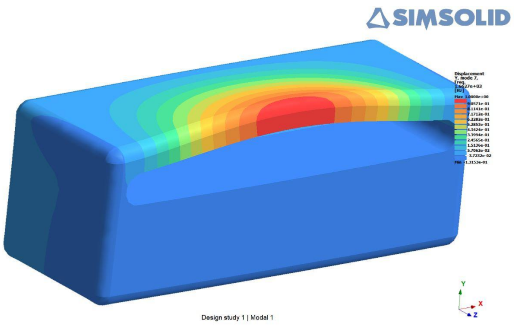


Figure 12. Preliminary simulation of the woodblock top plate first mode vibrations, with SimSolid software.

Also, the matching of the two lower resonance peaks, as shown in Figure 8, probably one from the Helmholtz and the other from the open-close pipe model, may also optimize the instrument's behavior.

It would be interesting to proceed with intensity and directivity measurements, such as using an acoustic camera to better map the sound diffusion patterns around the woodblock; however, from our perception and experience with the instrument, we expect much higher sound emission from the slit, certainly the main sound diffuser.

The woodblock base must be sufficiently thick to provide the instrument with the mass ground, but the dimensions are not critical as in most idiophones, such as xylophones. In principle, due to their acoustical natures, a vibration transducer, such as an accelerometer, would efficiently pick up the body vibrations of an aerophone, and its sound diffusion to the ambient would mostly occur around the solid oscillating shell. In aerophones, the main sound emission will take place across the open-air holes and bells, preferably transduced by a microphone. These contrasting methods of energy conversion, according to their results, might serve as a technical way of deciding if the instrument is an idiophone or an aerophone, or both.

This study is confined to a single woodblock. However, we are aware of the fact that a larger variety of samples will better represent this particular instrument, also improving the algorithm for the prediction of its resonances. We plan to include more blocks in the continuation of this project.

7. CONCLUSIONS

The woodblock is usually described as a very simple instrument, but we should recognize that there are several interesting and complex issues regarding its sound production and corresponding perception.

The calculations of natural frequencies of cavities, such as the woodblock slit, need a better and more accurate model, for the Helmholtz and the "pipes" resonances, but the estimates were rather close to experimental data, with an error usually lower than 10%. This means that our model, if improved and provided with factors for length, mass and volume correction, may yield a more accurate algorithm for the actual resonance peaks.

We should aim at a complete algorithm taking into account the expected pitches, the chosen materials and the woodblock overall dimensions, using a combination of methods, including finite elements for the idiophonic aspects and air cavity and air column calculations for the aerophonic response.

This study may be applicable to the optimization of woodblocks and similar instruments, such as temple blocks and slit drums [19]. This approach may assist in the project and production of more affordable instruments, with recycled wood or other low-cost materials.

There remains no doubt that the woodblock, if properly built, is a pitched instrument; however, it may often present sounds with two or more different pitches.

It was noticed that partially covering the slit opening causes the pitch to drop. This seems reasonable if we regard the slit cavity as a Helmholtz resonator where the "neck" area is reduced. Also, we observed that this area reduction corresponds to a reduction in the first resonance, i.e., the 550 Hz peak in Figure 6. The other peak (ca. 640 Hz) remained unchanged. This indicates that the first resonance is related to the slit cavity, while the second resonance depends on the top plate properties.

This investigation leads to potential perspectives in the performance techniques of woodblocks. For instance, the placement of the hand or the forearm along the slit opening allows the control of the instrument's pitch and intensity, as shown in the following video [21].

Also, we may design more efficient woodblocks, with positive environmental impact through the use of recycled and more affordable materials, including the replacement of traditional and sometimes endangered woods with plastic, resinous polymers, and others.

We may propose the concept of a "pure idiophone", as an instrument that presents a satisfactory and representative vibratory behavior, even in the case of a complete absence of air. For instance, a cymbal or a rattle in the vacuum, if properly supplied with accelerometers, will provide a signal that is considered similar to the one picked by a microphone, under normal conditions. On the other hand, a woodblock, a temple block, an udu and a few additional instruments, would generate a very different signal in the vacuum.

The woodblock, containing a Helmholtz resonator, may therefore be regarded as a Helmholtzian instrument [22].

Also, new designs of the woodblock may be devised, with ergonomic features and practical devices that facilitate the change in the slit dimensions while playing. From our data, observations, and discussion, it seems acceptable and appropriate to refer to the woodblock as both an idiophone and an aerophone.

Acknowledgments

The author is indebted to Dr. Sérgio Naidin, percussionist and ethnomusicologist, who helped with the provision of valuable information and generously made available several woodblocks for the experiments and recordings. Professor Fernando Castro Pinto and Tiago Paulino must be acknowledged for the preliminary simulations and vibrational measurements at LAVI-UFRJ. Also, the discussions with percussionists and professors Eduardo Tullio, Eduardo Pacheco, and Leonardo Gorosito were very much appreciated.

8. REFERENCES

[1] M. I. Okos. Use Of Woodblock as Instructional Aid in Teaching Simple Rhythm. The Pedagogue: Festschrift In Honour Of Professor Chukwuemeka Eleazar Mbanugo, v. 1, n. 1, 2021.

[2] B. Hopkin. Musical Instrument Design: Practical Information for Instrument Design, Sharp Press, 1996

[3] E. C. Udok, E. Ukamaka Ibekwe. “The Slit Wooden Drum In Southern Region Of Nigeria: Musical And Extra-Musical Values”. Awka Journal of Research in Music and the Arts (AJRMA) Vol. 14 2020. pp.273-290

[4] D. R. Gilbert. “The Lukumbi; a six-toned slit drum of the Batetela”, African Music Society Journal Vol 1, number 2 (1955) pp. 21-23

[5] A. Cannaday, T. Moore. “Tuning parameters of the Nigerian slit drum”. The Journal of the Acoustical Society of America 130, 2508 (2011); https://doi.org/10.1121/1.3654993

[6] L. Reymore, D. Huron. “Identifying the Perceptual Dimensions of Musical Instrument Timbre”, Proceedings of ICMPC15/ESCOM10. Graz, Austria, 2018, pp. 372-377

[7] V.-C. Mahillon. Catalogue descriptif et analytique du Musée Instrumental du Conservatoire Royal de Musique de Bruxelles, Vol. I, 2nd ed., Ghent, 1893

[8] E.M. Von Hornbostel, C. Sachs. Classification of musical instruments: Translated from the original German by Anthony Baines and Klaus P. Wachsmann. The Galpin Society Journal, p. 3-29, 1961.

[9] C. Sachs. The History of Musical Instruments, New York: W. W. Norton, 1940.

[10] L. Eargle. Music, Sound, and Technology. 2nd ed. ISBN 978-1-4757-5938-9, 1990

[11] F. Crane. “The Jew's Harp as Aerophone”, The Galpin Society Journal, Vol. 21 (Mar., 1968), pp. 66-69

[12] B. Anderson, C. B. Hilton, F. Giorgini. “Equivalent circuit modeling and vibrometry measurements of the Nigerian-origin Udu Utar drum”, The Journal of the Acoustical Society of America 133, 1718 (2013); doi: 10.1121/1.4789892

[13] N.H. Fletcher, T. Rossing. The Physics of Musical Instruments. New York: Springer.1998.

[14] B.H Repp. The sound of two hands clapping: An exploratory study. The Journal of the Acoustical Society of America, 81(4), pp.1100-1109, 1987..

[15] D. Pantalony. Altered Sensations: Rudolph Koenig's Acoustical Workshop in Nineteenth-Century Paris. Dordrecht: Springer, 2009.

[16] V. Bukur. Handbook of Materials for Percussion Musical Instruments, https://doi.org/10.1007/978-3-030-98650-6_12, 2022

[17] R. Hartenberger (Editor). The Cambridge Companion to Percussion, Cambridge University Press, 2016

[18] A. Schneider, K. Frieler. “Perception of Harmonic and Inharmonic Sounds: Results from Ear Models”. Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music. CMMR 2008. Lecture Notes in Computer Science, vol 5493. Springer, Berlin, (2009)

[19] M. McPherson, J. McDermott. ”Diversity in pitch perception revealed by task dependence”. Nat Hum Behav. 2018 January; 2(1): 52–66, doi:10.1038/s41562-017-0261-8

[20] J. Blades, J. Holland."Woodblock". The New Grove Dictionary of Music and Musicians, second edition, 2001.

[21] L. Fuks. A short video demonstrating a technique for controlling the pitch of #7 Ron Vaughn Woodblock. The forearm changes the slit opening area, therefore changing the resulting pitch (2023). https://youtube.com/shorts/MXN52TxVs80?feature=share

[22] L. Fuks, T. Grothe, M. Kob. Development and demonstration of Helmholtzian instruments, Proceedings of DAGA 2021, Wien, 2021, pp. 463-465.

MIPCAT - A Music Instrument Performance Capture and Analysis Toolbox

Andre Almeida
UNSW Sydney
a.almeida@unsw.edu.au

Weicong Li
Western Sydney University
weicong.li@westernsydney.edu.au

Emery Schubert
UNSW Sydney
e.schubert@unsw.edu.au

Joe Wolfe
UNSW Sydney
j.wolfe@unsw.edu.au

John Smith
UNSW Sydney
john.smith@unsw.edu.au

ABSTRACT

Playing a musical instrument to convey a convincing and engaging performance requires mastering several musical and technical aspects of playing the instrument. Timing and loudness of notes are recognised as important components of conveying musical expression, but also important are finer aspects such as the timbre of notes, how rapidly a note starts as well as fine variations of loudness and pitch within the note. An expert musician acquires a subtle control (often subconscious) of the gestures needed to produce these sound results, but they are usually difficult to observe or communicate. This presentation introduces the MIPCAT, a hardware and software toolbox that can be used to record simultaneously the sound and the action of a musician: the variables that directly affect the sound such as (in the example of a reed instrument) blowing pressure and bite force, but also body gestures or mouth-mouthpiece geometry, captured via general-purpose cameras. The toolbox also facilitates data processing and analysis in a semi-automatic way. To demonstrate the potential use of the MIPCAT in pedagogy, we show measurements of the gestures of a beginner clarinettist in comparison with those of a panel of expert players.

1. INTRODUCTION

Western music performance traditions involve more than just converting musical notation (the 'score') to a series of note pitches and durations. The score can include what is called expressive indications, telling the musician how loud to play, how loudness should change in a particular stretch of the score, or how the duration of the tone should change relative to the indication given by the note figure in the score.

All this information printed in the musical score still leaves a margin of liberty of individual expressiveness to the musician. In some instruments, this liberty is limited to slight changes in the duration and timing of the note: for example, in baroque keyboard instruments such as the organ or the harpsichord, expressiveness is mostly a fine art of adjusting the durations of a note to the musical context inside a phrase or within the piece, or varying the relative start times of notes in a chord (written as simultaneous). In a piano, the range of available parameters increases, as the loudness can also be changed, or the ringing of the notes with the sustain pedal. In many instruments, such as most of the winds, the range of parameters is greatly increased because the control by the musician for each note is exerted throughout the duration of the note, allowing for changes of the note envelope, multiple aspects of its timbre, and slight modifications of the pitch.

In the clarinet, for example, two important parameters are blowing pressure and reed bite force. Not all possible values can be used for these two parameters: there is a limited range that allows for the production of a periodic tone. Within this range, only a smaller range has "aesthetically suitable" applications and an even smaller range will be used by a particular player, in a particular musical context.

To play a tune requires a player to negotiate a path through a limited volume of the space of control parameters so that the notes sound, the slurs are smooth and the timbre homogeneous. This was the object of a challenge that our team participated in, successfully providing a mapping between player parameters and musical pitches so that a robot musician could play most of the range of the clarinet, for real musical pieces (http://www.phys.unsw.edu.au/jw/clarinetrobot.html).

On the other hand, using a fixed pair of values for a given note results in a rather mechanical performance. Expressivity and "humanness" are achieved by exploring the range of aesthetically sound parameters in a performance. Musicians practice for years in order to acquire an intuitive sense of the right chaining of parameters in a note, in a phrase and an entire musical piece.

Several authors have focused on the slight variations in the characteristics of the sound, in particular tempo and loudness, either due to some implicit rules given the structure of the score, to convey some kind of emotional intention by the performer, or just to personalize the rendition of a musical piece. Timing and loudness aspects have been covered widely in the literature [6]–[8], presumably because those are aspects that are common to many instruments. Other aspects such as intonation and timbre variations have also been studied, although less frequently [9]. Much less focus has been given to how the adjustments to the sound results are achieved by musicians, but see work from the Vienna group [10].

Relatively little is known about the reasons behind the set of parameters chosen by a musician to play a particular note with a particular sound result: Is this an individual choice and can different player parameters be used to achieve similar sound results? Does this choice depend on the musical context or the expressive intention? Can the knowledge of possible parameters be used to teach how to play a musical instrument? Musicians can describe the actions they need to undertake to modify the emotion conveyed by a performance [11], but how do these intentions correlate with their actual, physical actions?

Some authors have incorporated sensors in musical instruments to study musical interpretation in physical terms [12]–[15].

The present work aims to provide a comprehensive and cost-effective way of acquiring and pre-processing the main musical and player parameters involved in the performance of a musical instrument. It is an extension of a reviewed article published recently [16]. In that paper, we foresaw applications in pedagogy. In this paper, we use the toolbox to compare measurements made on a beginner clarinettist with a collection of measurements made on experienced professionals.

The toolbox is here applied to the clarinet, but with some effort, it could be adapted to other wind instruments. For other families of instruments, such as bowed strings, for instance, some of the software tools may be applied and the automated tracking of ArUco tags could also be adapted to the motion of the bow. Different sensors would be used, for example force sensors in the bridge and bow and accelerometers in the latter.

2. THE MIPCAT

2.1 Hardware

The capture system of the Music Instrument Performance Capture and Analysis Toolbox (MIPCAT) consists of several sensors fitted to a clarinet mouthpiece (Figure 1), two microphones capturing the external sound, and three cameras.

The sensors were:

- A tonguing sensor consisting of a small wire glued onto a *Légère* reed. A second wire was connected to the thumb rest on the clarinet so that when the tongue touched the wire, an imperceptible electric current (of a few μA) would flow through the body of the musician;
- An optical sensor measuring the distance between the mouthpiece and the reed, by sensing differences in reflected infrared light on the reed;
- A miniature pressure sensor measuring pressure (DC and AC) inside the mouth of the musician;
- A second miniature sensor measuring the DC and AC pressure inside the mouthpiece;
- A B&K microphone measuring AC pressure fluctuations inside the barrel.

Apart from the B&K microphones and the external audio microphone, all the sensors were plugged into a custom-built conditioning system, whose schematics can be found with the software package. All of the electronics, including proprietary apparatus, the acquisition system and the laptop were powered by a 12V car battery, to ensure the electrical safety of the musician and experimenter.

Of the two external microphones, an audio microphone was placed on a stand about 50 cm from the player. A second microphone was a B&K measurement microphone attached to the lower part of the instrument, below the right hand.

Two cameras captured a front and a right-hand side view of the clarinet player and the instrument (Figure 2). Four ArUco tags (QR-code style markers [17]) were attached to the clarinet. These are easy to detect with image-processing tools. A third miniature camera was attached to the barrel of the instrument and captured a side view of the mouthpiece and the player's mouth (Figure 3). With a coloured tag and a scale glued to the mouthpiece, this allowed the position of the lips along the mouthpiece of the instrument to be measured.

The hardware components are specific to the clarinet, but with small modifications, the sensors could be adapted to other single-reed instruments such as a saxophone. Double-reeds or flutes would require greater modifications, but solutions have been found to measure mouth pressure in flutes [13] or brass instruments [18], [19].

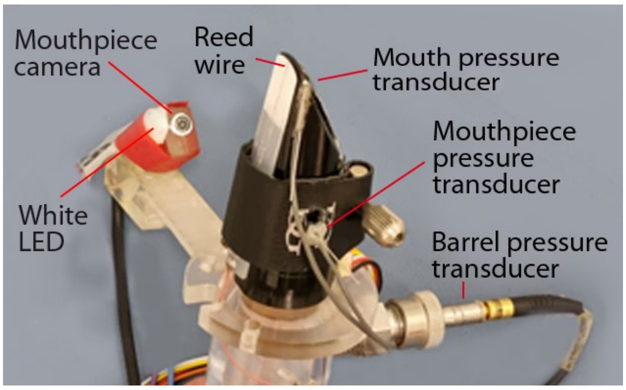


Figure 1. The sensor-fitted mouthpiece used in the MIPCAT (from [16])

Figure 2. One of the authors demonstrating the use of the sensor-fitted clarinet, and the use of ArUco markers for clarinet position tracking (from [16]).

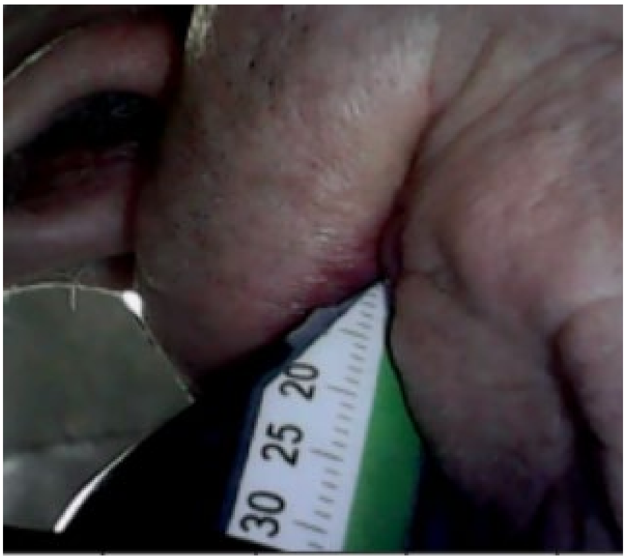


Figure 3. The mouthpiece fitted with a coloured scale used for tracking the lip position on the mouthpiece (from [16]).

2.2 Software

A software package, mostly written in Python, is publicly provided to pre-process and analyse these data and comparable sets. This package is available on GitHub at https://github.com/goiosunsw/mipcat. There are several components to this package:

2.2.1 Video to Audio Alignment
This module calculates the delay between a reference signal, for instance, that of the external microphone, to the audio track associated with a recorded video. The alignment is based on a fingerprinting algorithm [20] that extracts relative peak positions in each of the signal's spectrograms.

2.2.2 Time-series Indicators
This module analyses the recorded data in frames, and calculates some indicators, in particular:

- DC offset: the mean value of the measured signal within that window;
- Amplitude: the RMS amplitude of the oscillation of the measured data;
- Frequency: the fundamental frequency of the oscillation in the measured signal;
- Harmonic components: amplitudes of five harmonics of the fundamental frequency.

The signal input to these analysers is first pre-processed to take into account any calibration needed by the signal. This and other signal properties are defined in a YAML file.

2.2.3 Segmentation and Alignment with the Musical Score
This module takes as input an audio recording of a performance (sometimes repeated several times) and a music score. Time instants in the audio signal are matched to note beginnings and endings.

2.2.4 Mouthpiece Video Processing
The scale glued onto the mouthpiece consists of a human-readable millimetre scale alongside a green strip that is easy to detect and isolate digitally using open-CV. The position of the numbers in the human-readable scale is tracked using a template tracker (based on correlation) for two reasons: firstly, it allows calibrating for physical distances, secondly, it indicates where the green strip should be found, making it easier to discard false detections of the green patch.

2.2.5 Clarinet Position Processing
This module uses open-CV to detect the position of the clarinet markers (ArUCo) on the image. Open-CV provides a tag detector for individual frames. A second layer of detection is added, consisting of a template tracker: a snapshot of the tag found with the standard detector is kept in memory, and a correlation algorithm tracks its position in a new image. This allows tracking of the tags even when they are partly obscured or blurred by motion.

2.2.6 Player Pose
The player is tracked using a deep neural network algorithm provided by Google and called Mediapipe. It can identify the position of key elements of the human body. Important for us are the position of the mouth and the head. This detection is not as accurate as the position of the clarinet based on the ArUco tags but can provide some rough indication, for example of the angle of the head.

2.2.7 Signal Collection and Building of a Database
Large volumes of data are generated by the descriptor extraction. The data set is easier to analyse by aggregating it within individual notes and calculating statistical values such as means, standard deviations, etc.

3. A WORKED EXAMPLE

3.1 Introduction

The primary purpose of the worked example was to use the MIPCAT system to demonstrate how player's physical gestures in a study concerning expressive playing are captured and processed. Detailed results from this study will be published later. Here we give an example of how a set of recordings from expert musicians can be compared to the performance of a beginner, pinpointing when the main differences arise and suggesting actions that are closer to the expert's technique.

The musical excerpt used in this example is an eight-bar section introducing the main theme of the slow movement of Mozart's Clarinet Concerto K. 622. Seven musicians were invited to participate in the study, 6 of them professionals playing in orchestras (used to obtain "reference expert performance"), and one beginner who has played other reed instruments for more than 30 years but never studied or regularly played the clarinet. They were involved in a 3-hour-long session (with a 30-minute break) concerning the expression of emotion through music, and a set of other tasks related to musical performance. As part of this study, they were asked to play the above-mentioned excerpt on the sensor-fitted clarinet (the lab clarinet) and also on their own clarinet.

3.2 The Recording Stage

The musicians sat in a low-reverberation room, as in Figure 2. They had some time to practise on the sensor-fitted instrument, and to select from a set of synthetic reeds according to which felt more comfortable to them. They played the excerpt twice on the sensor-fitted instrument and were then asked to play it again on their own instrument. If needed they could reject a recording and repeat it. An experimenter from the team provided guidance and launched the recording of the sensor data, together with each of the three cameras.

3.3 Pre-processing and Segmentation of the Audio

Once the data have been collected, the first step is to identify the individual notes, matching them to the score. This can be done manually with audio tagging software, but MIPCAT provides a set of tools that can help with this task.

Initially, a series of descriptors are extracted from the audio (Figure 4). Among these are the amplitude and frequency of the sound, but many others are extracted also from the sensor recordings, to be used at a later stage. Extraction of the descriptors can be performed automatically for a single recording or a set of recordings using the script "ts_gen_from_csv".

Amplitude and frequency from a reference channel (usually the internal instrument pressure if available, otherwise from one of the external microphones) are used to detect note transitions, in a first pass, and then to align these note transitions to the score. These two passes can be performed in a single step using the script "note_matcher". Once again, the script can be run individually for one recording, or a set of recordings. The script outputs a TextGrid file per recording, and these files can be used in *Praat* (https://www.fon.hum.uva.nl/praat/) to check and adjust the segmentation.

3.4 Pre-processing of Video Recordings

MIPCAT provides a set of tools to help extract key positions in the video files. Some of these are specific to our project, but the ArUco tagging is quite generic and can be used in many different situations. The script that tracks the ArUco markers is more robust than the simple framewise detection of a tag, using information from previous frames to keep identifying the tag even if it is blurred by motion or partially obscured. Tracking can be performed with the script "aruco_tracker".

Tracking of human position can also be used in many different situations and is done in an automated way by the script "mp_pose_detect" which calls Mediapipe in an automated way and exports the data to be processed later. Finally, the mouthpiece video is specific to clarinets and saxophones. This stage has to be run for each video in two stages: first calling the GUI "mouthpiece_gui" to manually adjust key parameters for the automated processing, and creating a configuration file for each video.

The tracking of the green patch that is partially and variably covered by the musician's lips is then done automatically with "mouthpiece_tracker".

3.5 Alignment of Video and Sensor Signals

The cameras used in this project were not synchronised to the sensor and microphone signals, because cameras that provide a synchronization signal can be expensive. Instead, the audio that is recorded by each camera together with the video is synchronised offline using a fingerprinting algorithm. This algorithm identifies key features in the spectrogram of each audio signal (camera and signal-synchronous) and runs a matching algorithm to identify the delay between the two signals. This is done with the script "align_keypoints.py"

Once the delays are known, the measurements extracted from the videos in the first step can be cut and aligned to the sensors so that they can be seamlessly analysed as the output of any other sensor.

3.6 Collecting Data

The data gathered with this system are quite comprehensive, and can be analysed in different ways. For the purpose of this worked example, we chose to aggregate signal data for each note, calculating average values, trends and variability measurements. For example, for the measured mouth pressure, many different indicators are calculated such as the average mouth pressure during a note, the standard deviation, the trend, the amplitude of the oscillation (since sound also propagates inside the mouth), etc.

These calculations are run for all the notes in all the recordings automatically with the script "build_note_database".

4. RESULTS

With the knowledge of the note boundaries for each performance, it is possible to align different performances on a modified time scale that is measured in musical beats instead of seconds. With this alignment, we can compare the descriptor time series for different players, as seen in Figure 4.

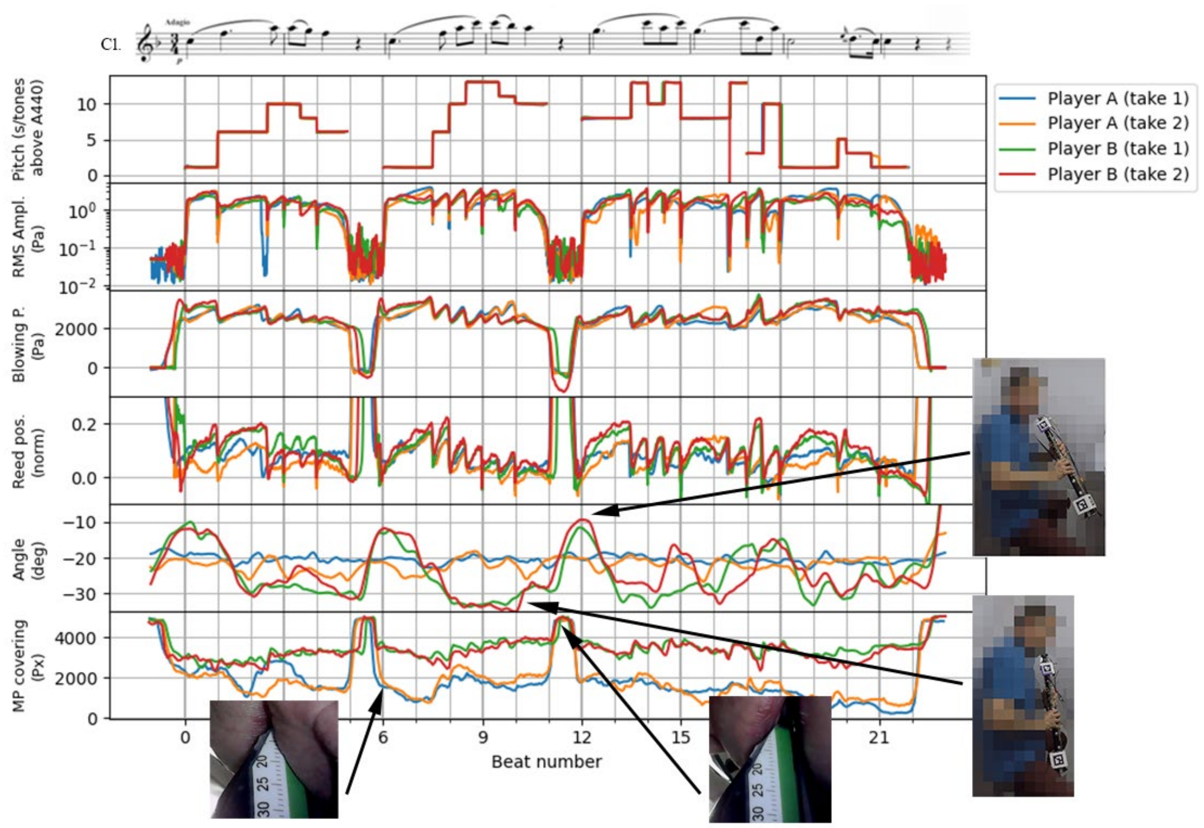


Figure 4. An example capture of descriptor time series corresponding to 4 performances of 2 (expert) players. From top to bottom, the score (transposed for the instrument), the playing frequency, sound RMS amplitude, blowing pressure, normalized distance of the reed to the lay of the mouthpiece, angle of the clarinet and covered area of the mouthpiece (from [16]). The photographs provide two extreme examples of mouthpiece covering (bottom) and clarinet angle (on the right-hand side).

4.1 Note-by-note Analysis

Instead of comparing "instantaneously" how a player is playing their instrument, we can analyse the performances in terms of averages of the descriptors within each note. In this way, it is possible, for example, to compare an individual performance with an "average performance" of a larger set of musicians that participate in a study. (Detailed analysis of a larger set is a project of the present team). In the following plots, we show the "average expert perfor mance" in blue, calculated as the median value of a particular descriptor for each note, for all the 6 expert players, and for both the lab instrument and their own instrument, except when the data involves player parameters such as blowing pressure or reed position. Overlapped is the variability of the descriptor, measured as the inter-quartile range for each note. In red we display the performance of a beginner, to compare it to the "average expert performance".

The first figure (Figure 5) shows the value of the note amplitude in a box plot without whiskers. The horizontal length of each box is the IOI for that note. Because each player can decide to play the excerpt with a different overall amplitude, the absolute amplitude for each note is subtracted from the average amplitude for the entire excerpt by that player. It thus shows the amplitude of the note relative to all other notes. This typically reduces the overall variability of each note's amplitude by 3 dB, i.e. instead of the amplitude of a typical note ranging from 97 to 102 dB (a range of 5 dB), the range is reduced to 2 dB.

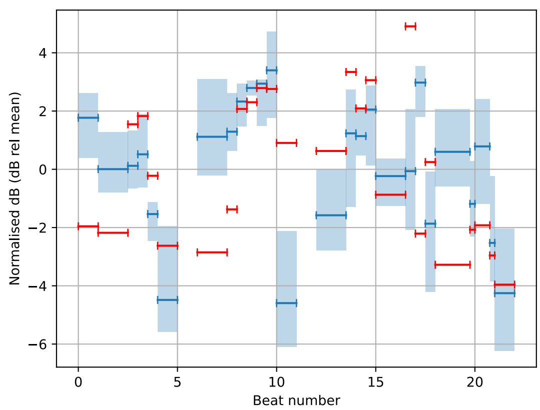


Figure 5. Box plot showing average (horizontal lines) and inter-quartile variability (shaded areas) of the sound amplitude of expert performances (blue), per note of the excerpt, compared to an excerpt from the beginner (red). Amplitude is shown as a difference from the overall excerpt amplitude.

The plot shows, for instance, that the beginner is exaggerating the dynamics at the beginning of the second phrase while not doing a large enough decrescendo towards the ends of the 1st and 2nd phrases.

Similarly, we can see in Figure 6 that the beginner is not coping well enough with the tuning in the higher notes. Notice that the variability in pitch across expert players is small in this plot.

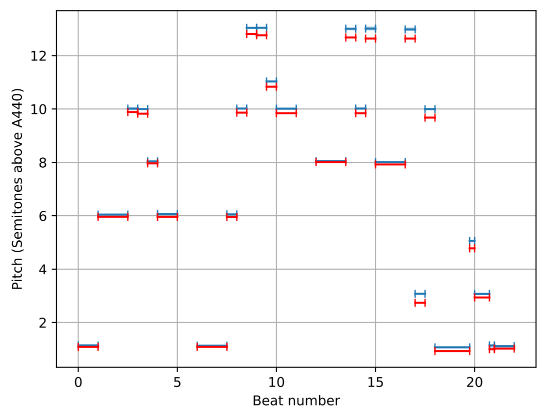


Figure 6. Average pitch of each note of the excerpt for expert musicians (blue) and a beginner (red). Unlike the other figures, the variability (inter-quartile range) is not visible because it has typical magnitude of a few cent (hundredths of a semitone).

The plots in Figures 5 and 6 would be possible to obtain using only an audio recording. The sensors attached to the clarinet however make it possible to look into the actions of the musician that is producing these sound results. Figures 7 to 11 show how it could be possible to pinpoint technical problems from a beginner

Figure 7 shows how the blowing pressure varies along the excerpt, for the "average expert" and for the beginner. For instance, the two high Cs immediately before and on the 9th beat, in the second phrase, were produced using considerably greater blowing pressure from the beginner (red dash) in comparison to the professional mean (blue).

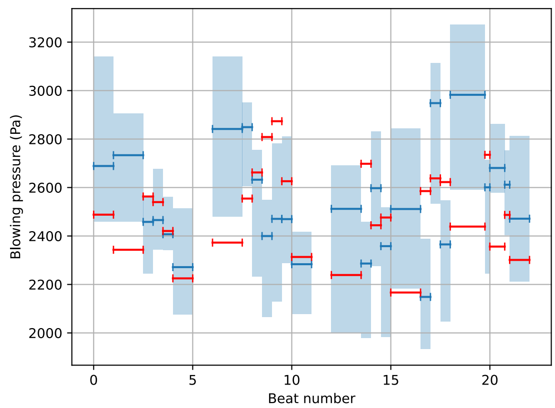


Figure 7. Average blowing pressure (blue) for expert musicians (per note) compared to a beginner's performance (red)

Another important parameter for clarinettists is the bite force applied to the reed. Our sensors do not measure force directly, instead measuring the reflectance of a light shining on the reed. This reflectance is a monotonic function of the distance, roughly linear in the range of interest. However, changing the reed, and to some extent changing the reed position on the lay may change the reflectance. For this reason, the data presented in Figure 8 is a difference in the reed position for the note from the average position for all notes in the recording. In this figure, it is apparent that the beginner changes the reed position more than the expert players.

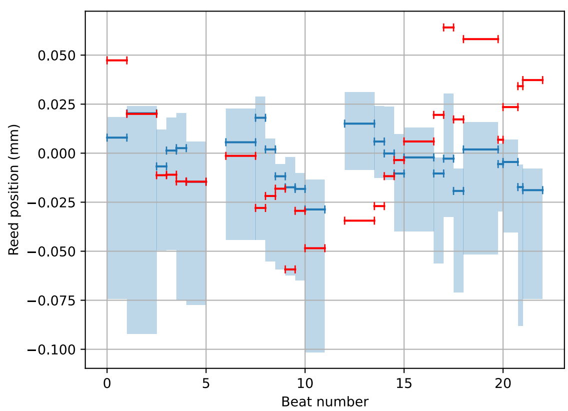


Figure 8. Average reed displacement from excerpt overall mean for expert musicians (blue) compared to a beginner (red).

The mouthpiece camera measured the amount of mouthpiece that is covered by the bite of the musician. It was found that different musicians can use rather different bite positions (or configurations, since the covered part of the mouthpiece measured in the image might not correspond exactly to the bite position), but the position is, on average, only changed by a small amount during playing (Figure 9).

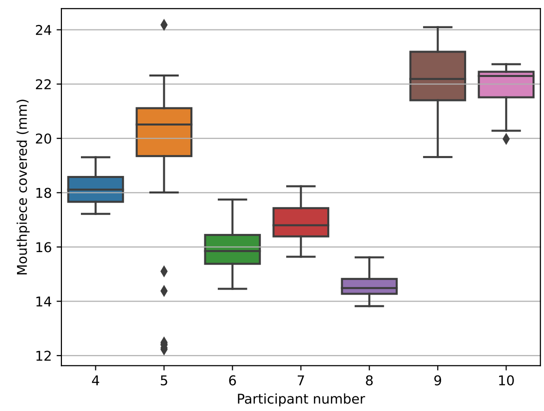


Figure 9. Average mouthpiece length covered as measured from the tip of the mouthpiece, for each of the professional (numbers 4-9) and beginner (number 10) participants.

Still, it is possible to observe a trend during an excerpt whereby the mouth slightly recedes from the mouthpiece in the first phrases (Figure 10). For the beginner the retraction is more pronounced towards the end of the excerpt.

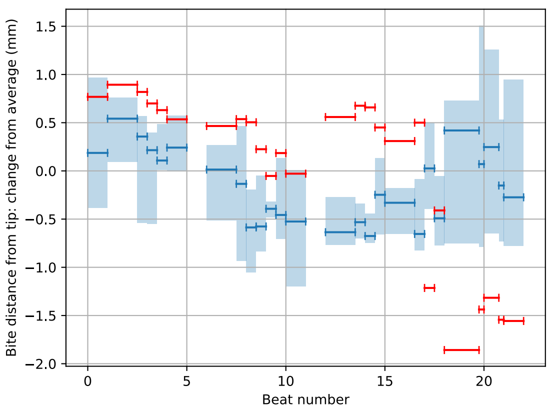


Figure 10. Difference between each note's bite position to the overall average position for each player. Blue: expert musicians, red: beginner.

The views of the player can be used to determine the clarinet's orientation relative to the musician's face. This is also a measure that can vary considerably among players as shown in Figure 11, not least because of physiological differences that may not translate directly into a different action on the instrument. It may also reflect different playing traditions and training.

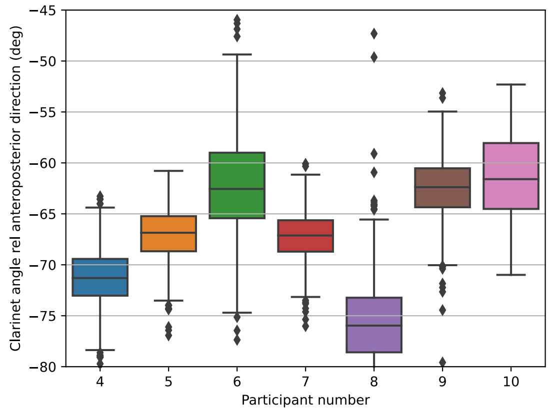


Figure 11. Average clarinet angle relative to forward facing direction of head for each player (4-9: professionals, 10: beginner). Notice how players 6, 8 and 10 move their instrument considerably more during the excerpt.

5. DISCUSSION AND CONCLUSIONS

Measurement systems for musical instruments and automated or semi-automated methods of processing data open new perspectives on how we teach and learn technical aspects of musical performance. Student performances can be compared to expert performances and technical styles can be compared. A range of other applications is foreseen in seeking a deeper understanding of how the musician-instrument combination works in vivo.

In the present study, a pedagogical application was shown, that can describe the range of playing actions of expert performances, against which that the student performances can be compared. This comparison goes beyond the simple auditory, interpretative level, but goes to the physical action level — the very act of playing. With some refinements, this has considerable applications for player training.

The software used for this analysis is available with documentation on https://github.com/goiosunsw/mipcat, and a larger dataset of performances will be made available shortly.

6. ACKNOWLEDGMENTS

We warmly thank the Australian Research Council for support of this project, Yamaha for the clarinet, Légère for the reeds and the participants for the study.

7. REFERENCES

[1] C. Fritz and J. Wolfe, "How do clarinet players adjust the resonances of their vocal tracts for different playing effects?," *J. Acoust. Soc. Am.*, vol. 118, no. 5, pp. 3306–3315, Nov. 2005, doi: 10.1121/1.2041287.

[2] A. Almeida, W. Li, J. R. Smith, and J. Wolfe, "The mechanism producing initial transients on the clarinet," *The Journal of the Acoustical Society of America*, vol. 142, no. 6, pp. 3376–3386, 2017.

[3] A. Friberg, R. Bresin, and J. Sundberg, "Overview of the KTH rule system for musical performance," *Advances in cognitive psychology*, vol. 2, no. 2, p. 145, 2006.

[4] A. Almeida, D. George, J. Smith, and J. Wolfe, "The clarinet: How blowing pressure, lip force, lip position and reed 'hardness' affect pitch, sound level, and spectrum," *Journal of the Acoustical Society of America*, vol. 134, no. 3, 2013, doi: 10.1121/1.4816538.

[5] D. Fabian, R. Timmers, and E. Schubert, *Expressiveness in music performance: Empirical approaches across styles and cultures*. Oxford University Press, USA, 2014.

[6] R. Bresin and G. U. Battel, "Articulation strategies in expressive piano performance," *Journal of New Music Research*, vol. 29, no. 3, pp. 211–224, 2000.

[7] W. Goebl, "Melody lead in piano performance: Expressive device or artifact?," *J. Acoust. Soc. Am.*, vol. 110, no. 1, pp. 563–572, Jul. 2001, doi: 10.1121/1.1376133.

[8] B. H. Repp, "Acoustics, perception, and production of *legato* articulation on a digital piano," *The Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3862–3874, Jun. 1995, doi: 10.1121/1.413065.

[9] M. Barthet, R. Kronland-Martinet, and S. Ystad, "Consistency of timbre patterns in expressive music performance," in *Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx-06), Montreal, Canada*, 2006.

[10] M. Pamiès-Vila, "Expressive performance on single-reed woodwind instruments: an experimental characterisation of articulatory actions," Doctoral dissertation, University of Music and Performing Arts, Vienna, 2021.

[11] A. Almeida, W. Li, E. Schubert, J. R. Smith, and J. Wolfe, "Expressive goals for performing musicians: The case of clarinettists," *Musicae Scientiae*, vol. 0, no. 0, 2022.

[12] V. Chatziioannou and A. Hofmann, "Physics-based analysis of articulatory player actions in single-reed woodwind instruments," *Acta Acustica united with Acustica*, vol. 101, no. 2, 2015, doi: 10.3813/AAA.918827.

[13] P. de la Cuadra, B. Fabre, N. Montgermont, and C. Chafe, "Analysis of flute control parameters: A comparison between a novice and an experienced flautist," *Acta Acustica united with Acustica*, vol. 94, no. 5, pp. 740–749, 2008.

[14] A. Hofmann, V. Chatziioannou, M. Weilguni, W. Goebl, and W. Kausel, "Measurement setup for articulatory transient differences in woodwind performance," presented at the Proceedings of Meetings on Acoustics, 2013, vol. 19. doi: 10.1121/1.4799441.

[15] W. Li, A. Almeida, J. Smith, and J. Wolfe, "How clarinettists articulate: The effect of blowing pressure and tonguing on initial and final transients," *Journal of the Acoustical Society of America*, vol. 139, no. 2, 2016, doi: 10.1121/1.4941660.

[16] A. Almeida, W. Li, E. Schubert, J. Smith, and J. Wolfe, "Recording and analysing physical control variables used in clarinet playing: A musical instrument performance capture and analysis toolbox (MIPCAT)," *Frontiers in Signal Processing*, vol. 3, 2023, Accessed: Feb. 19, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frsip.2023.1089366

[17] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, 2014, doi: 10.1016/j.patcog.2014.01.005.

[18] H. Boutin, N. Fletcher, J. Smith, and J. Wolfe, "Relationships between pressure, flow, lip motion, and upstream and downstream impedances for the trombone," *The Journal of the Acoustical Society of America*, vol. 137, no. 3, pp. 1195–1209, 2015.

[19] V. Fréour and G. P. Scavone, "Acoustical interaction between vibrating lips, downstream air column, and upstream airways in trombone performance," *The Journal of the Acoustical Society of America*, vol. 134, no. 5, 2013, doi: 10.1121/1.4823847.

[20] P. Cano, E. Batlle, E. Gómez, L. de C.T.Gomes, and M. Bonnet, "Audio Fingerprinting: Concepts And Applications," in *Computational Intelligence for Modelling and Prediction*, S. K. Halgamuge and L. Wang, Eds. Berlin, Heidelberg: Springer, 2005, pp. 233–245. doi: 10.1007/10966518_17.

How a Clarinettist Conveys Emotion in Music Playing: Measuring Player Gestures and Signal Parameters Using a New Toolbox

Weicong Li weicongli@outlook.com

Andre Almeida a.almeida@unsw.edu.au

Emery Schubert e.schubert@unsw.edu.au

John Smith john.smith@unsw.edu.au

Joe Wolfe j.wolfe@unsw.edu.au

UNSW Sydney

ABSTRACT

Instrumental music has the fascinating capacity to communicate emotions, sometimes with subtlety. Most previous research has focused on comparing features of recordings (e.g. sound level, tempo and timbre) of music conveying different emotions, rather than detailed gestures that musicians use during music playing, which would be useful for music students, teachers and music researchers. This paper reports an experimental study of how an expert clarinet player expresses three different emotions when playing the same pieces of music: *happy*, *sad* and *lacklustre/deadpan*. Parameters showing the musician's continuous control of blowing pressure and reed position were measured, as well as variables mentioned above in the recorded music; all were analysed semi-automatically using a new toolbox developed for this study. The results show how the emotions can be differentiated not only by the musical feature variables, but also the details of how the musician physically controls the instrument to produce them. These results provide an alternative approach for training musicians about expressive playing.

1. INTRODUCTION

Conveying various types of emotion plays an important role in music playing and this often requires that musicians master control over sets of playing techniques to physically manipulate the musical instrument to achieve desired musical goals. Good players can perform the same piece in different ways to transmit distinctly different emotions or expressive goals (EGs). However, the musical parameters used to produce EGs are better understood than the physical gestures used to produce them these musical parameters. Different EGs are conveyed in the presence of different musical parameters such as note length, loudness, pitch and timbre (see, for example, [1, 2]). The musical parameters that convey different EGs in performance have been studied over at least quarter of a century [1-4]. For example, Gabrielsson and Lindström [5] summarised a lexicon of musical elements/parameters that are associated most often with particular EGs.

These musical features are necessarily created by carefully controlled physical gestures or 'playing control parameters' by the musician. Understanding both the musical and gestural/control parameter aspects of performance is currently a gap in research on music expression, and closing this gap would have several applications. Such knowledge would provide insight into the nexus between the physics and the psychology of music performance, and would be useful for music students, teachers and music researchers who seek to understand music performance in terms of the physical control of the instrument, rather than, or in addition to, the more commonly used musical parameters (e.g. 'music systemisers' could benefit from such researcher – Kreutz [6]). Measuring parameters of both aspects simultaneously, and understanding how they are varied to convey different EGs is the aim of this pilot study.

Less attention has been given to the playing control parameters (hereafter called playing parameters) because they are difficult to measure, and measuring physical parameters of the instrument-player interaction usually requires invasion of the instrument and the player by measurement tools, which can interfere with the playing itself. Nevertheless, in the case of the clarinet, a few playing parameters used by players to control the sound have been studied, such as the average or DC air pressure in the player's mouth while blowing the clarinet, the position and vibration of the reed, the lip and tongue action, and some information about the acoustics of the player's vocal tract [7-13]. These studies used either single notes or simple excerpts with the focus of understanding the relation between player's input gestures and output sound, thus little attention was given to the particular EGs (if any) that were used in performance. More recently (and while many laboratory activities were limited by the COVID-19 pandemic), we used a survey to study how experienced clarinettists think they would play in order to distinguish different EGs in terms of both musical and playing parameters when performing the same musical excerpts [14]. Based on similarities, a set of six clusters of EGs were suggested for use in future studies. A few musical and playing parameters were reported by clarinettists as important in achieving specific EGs.

Over the last decade or so, we developed a musical instrument performance capture and analysis toolbox (MIPCAT) to capture and study various musical and playing variables controlled by clarinettists while performing music [15]. The toolbox includes both hardware and software. The hardware consists of various sensors mounted in or on the clarinet, plus microphones and video cameras. These can capture the player's blowing pressure, the sound pressure in the mouth and the instrument, the reed position and vibration, aspects of embouchure including the bite on the reed, some motions of the player's body and the output sound at different positions. The software contains several tools to process and analyse recordings of player performances captured by the hardware semi-automatically. Among other applications, the components of the toolbox enable us to study how clarinettists play the same music to convey different EGs, from a rich set of data that include both musical and playing parameters.

As an exploratory investigation for a larger study, this paper reports a case study of how a clarinettist plays the same musical excerpt with three different EGs and compares the musical parameters produced and some of the playing parameters used in producing them.

2. MATERIALS AND METHODS

In this study, MIPCAT and a modified clarinet (Yamaha YCL250 model with Yamaha 4C mouthpiece) were used for data acquisition and processing: see Figure 1. More details of the setup are described in [16].

One clarinettist having extensive classical and jazz playing experience participated in this study. From written music provided, the participant was asked to play the excerpt 'Happy Birthday to You' in G major (music score shown in Figure 2). The tune was played to convey three different EGs: *happy*, *sad* and *lacklustre/deadpan* (see [14] for detailed discussions of these EGs). Before recording, the participant was allowed to practise on the instrument modified to fit MIPCAT until feeling comfortable with it. The music and instructions were emailed to the participant beforehand so that he or she had time to prepare, if needed. To convey each EG, the participant was instructed not to change the melody but to vary freely other aspects of the performance, including tempo, dynamics, articulation, and timbre, to communicate the intended EG to listeners as convincingly as possible.

The following signals were used in the analysis:

- Mouth pressure: measured by a miniature pressure sensor (8507C-2, Endevco, Irvine, CA) fitted into the corner of the mouthpiece with its sensing membrane exposed to the inside of the player's mouth during playing;
- Reed position: the AC and DC components of the displacement of the reed in a direction at right angles to the instrument axis were measured by a reflective, infrared proximity sensor (QRE1113, ON Semiconductor, Phoenix, AZ) mounted inside the mouthpiece, 5 mm from the mouthpiece tip, directly opposite the reed;
- Radiated sound: measured by a ¾" microphone (RODE NT3, Sydney, Australia) mounted on a stand at the same height as the bell and at a distance of 45 cm. Microphones were also mounted on the clarinet bell and barrel, to separate out the effects of player and instrument motion. The recordings from these last two are not used in the current preliminary investigation. Measurements were conducted in a room treated to have low reverberation. More details about how the signals were acquired and processed were described in [15].

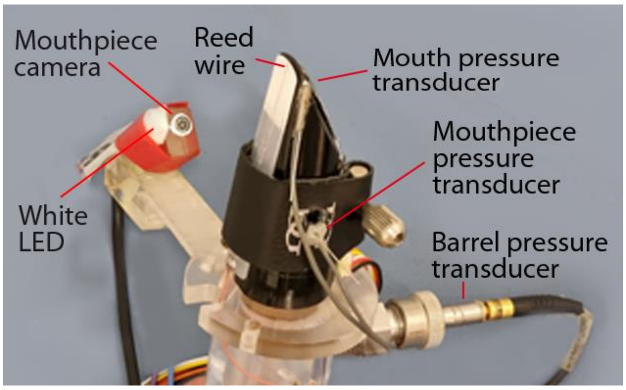


Figure 1. The mouthpiece and some of the sensors used in the MIPCAT. Reproduced from [15].

The software tools of MIPCAT were used for data processing. First, several indicator timeseries such as DC values, amplitudes of oscillation, fundamental frequency, etc. were extracted from the raw data. Then these timeseries were segmented note by note by matching the fundamental frequency measured to that of the music. Then some of the musical and playing parameters were extracted and averaged within each note from the timeseries: RMS sound pressure level, tempo, spectral centroid, blowing pressure and DC reed displacement. A local 'tempo' for each note was calculated by the written duration of the note divided by the Inter-Onset Interval (IOI) in minutes. IOI is the time interval between the onsets of successive notes.

Two takes from the participant gave a total of 50 tokens (notes) for each EG, resulting a total of 150 tokens. In addition to MIPCAT, MATLAB and R were also used for subsequent analysis.

3. RESULTS AND DISCUSSION

3.1 Musical Parameters

In this pilot study, we only explored a few of the musical parameters: sound pressure level, tempo and spectral centroid. These were considered important musical parameters that can characterise the performances and reflect an important subset of the musical parameters used to achieve a particular EG; we discuss them first before discussing control parameters.

Figure 2 shows the average sound level of the radiated sound for each note and the standard deviation of level within that note, for each of the three EGs: *happy*, *sad* and *lacklustre/deadpan*. (The last of these is hereafter abbreviated as *deadpan*.) In general, notes played to convey *happy* have the highest average sound level and largest standard deviation of level within each note. This result is consistent with previous research, e.g. [17]. Those for *deadpan* show medium average sound level and smallest standard deviation, and those for *sad* show lowest average sound level (also consistent with the literature) but

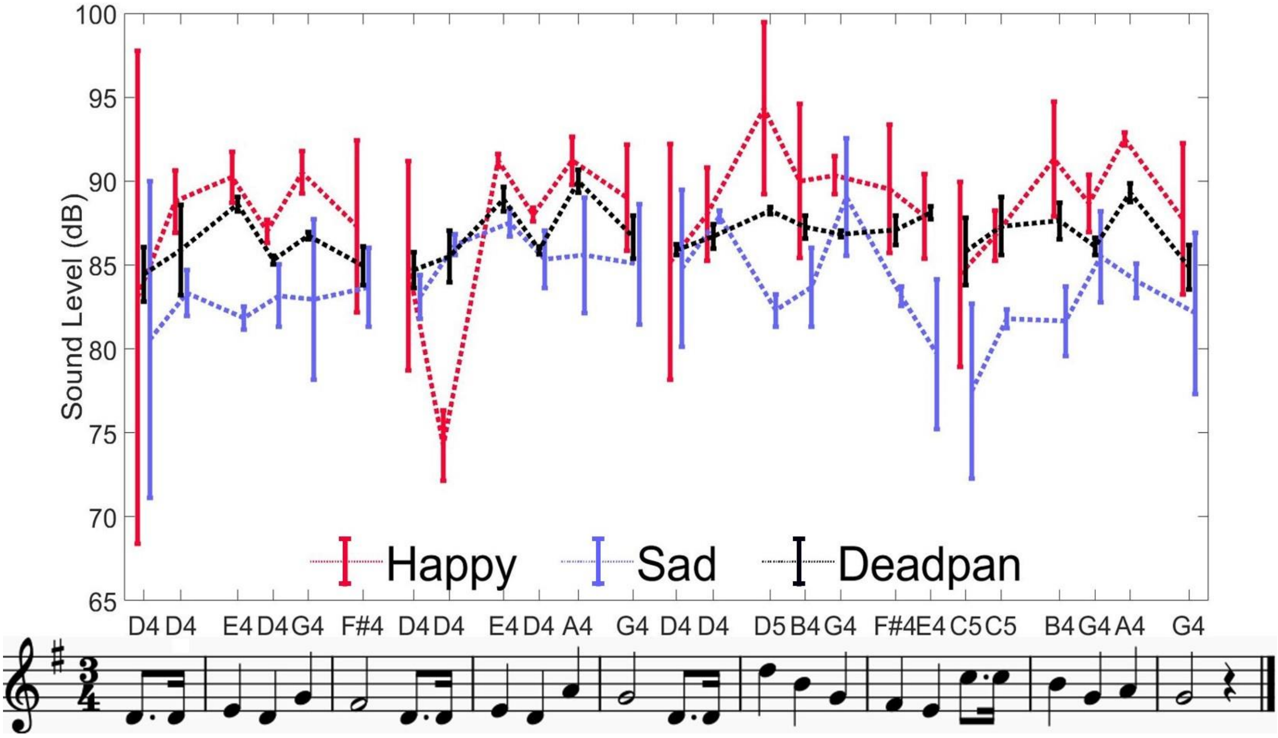


Figure 2. Sound pressure level of the radiated sound for each note and standard deviation of level within that note for the three expressive goals: *happy*, *sad* and *deadpan*. Notes of each phrase are connected with dotted lines. The pitches from the score (Happy Birthday to You) are shown on the x-axis, in order but not proportional to time.

also a large standard deviation within notes. In addition, notes at the start and end of each phrase often show larger variation than other notes in the phrase, e.g. notes D4 and F#4 in the first phrase for *happy*. Another observation is that different EGs show different patterns in the average sound level and standard deviation of the note sequence, e.g. *happy* and *deadpan* have similar patterns for the first, second and last phrases, whereas *sad* has a very different pattern from *happy*: e.g. the variation between notes in the first two phrases are smaller, but much larger in the last two phrases; the peak sound level of the third phrase falls on the third last note (G4) instead of the third note (D5, highest note in that phrase); the average sound level of the last phrase is significantly lower than the other three phrases. These features seem to indicate a particular playing style of the participant when conveying *sad* in the excerpt.

Figure 3 is the boxplot of sound levels of all the notes from the radiated sound for the three EGs. This figure further confirmed the observations from Figure 2: median sound level of *happy* is 88.4 dB, substantially higher than *deadpan* (86.8 dB) and *sad* (84.7 dB). As expected, *happy* and *sad* show more variation in the sound level than *deadpan*.

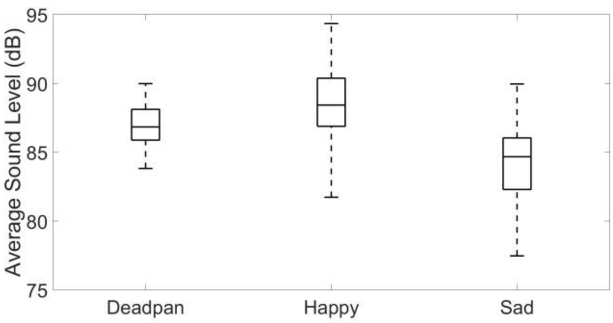


Figure 3. Boxplot showing the note-to-note variations in sound pressure level of the radiated sound for the three expressive goals: *happy*, *sad* and *deadpan*.

Figure 4 is the boxplot of local ‘tempos’ (in the unit of beats per minute) of all the notes, showing the note-to-note variation for the three EGs. *Sad* has the slowest median tempo value (107 bpm) but largest note-to-note variation in the tempo among the three EGs. *Happy* has median tempo (120 bpm), similar to *deadpan* (121 bpm), but much larger variation in local tempo than *deadpan*. The low variation in sound level and tempo for *deadpan* were expected. But it is interesting to note the similarity in tempo between *happy* and *deadpan*.

Figure 4. Boxplot showing the note-to-note variations in local tempo for the three expressive goals: *happy*, *sad* and *deadpan*. Local 'tempo' of each note was calculated as the written value of the note in beats divided by the Inter-Onset Interval in minutes.

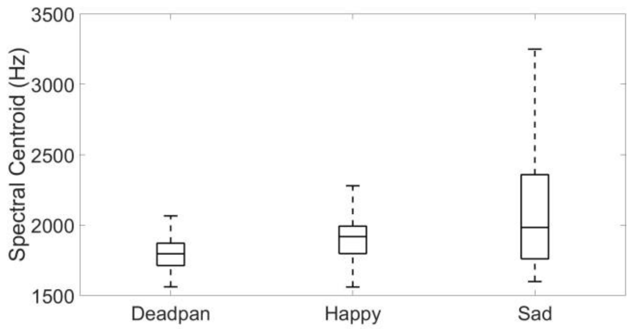


Figure 5. Boxplot of spectral centroids of all the notes for the three expressive goals: *happy*, *sad* and *deadpan*.

Figure 5 is the boxplot of spectral centroids of all the notes for the three EGs. The spectral centroid can be pictured as the 'centre of gravity' of the pressure spectrum. It is strongly correlated with perceived brightness of timbre [18]. Among the three EGs, *deadpan* has the lowest median spectral centroid at 1796 Hz and smallest variation in the spectral centroid, followed by *happy* (1918 Hz) and *sad* (1983 Hz). This finding is somewhat consistent with the literature. For example, one study found that spectral centroid for depressive sadness was lower than for happiness [19]. We can reconcile the difference between that study and ours by the stimuli used. The previous study [19] required rating of emotions of single tones played across a range of musical instrument timbres. This meant that those participants had far fewer cues from which to determine differences in emotions (a single pitch, controlled volume etc.), making participants rely more on timbral cues, and hence leading to more distinct results than the present study, where more musical context and flexibility allowed for several other cues to distinguish EGs, and where changes in timbre were 'within-instrument' changes, and therefore generally quite subtle. However, *sad* also has a considerably wider range of spectral centroids employed, suggesting that the player is varying timbre more in comparison to *happy* and *deadpan* to achieve the goal of *sad*. The spectral centroid may be expected to be influenced by blowing pressure and reed position [9]: greater blowing pressure and smaller average reed displacement from the mouthpiece both enhance clipping, which increases high harmonics and thus spectral centroid. Here, the written D4 notes are selected for calculating the correlation with blowing pressure and reed displacement: both correlation coefficients are non-significant ($r = 0.05$, $p = .75$ for spectral centroid and blowing pressure; $r = 0.14$, $p = .35$ for spectral centroid and blowing pressure). The reason of the non-significant result may be that the player was adjusting both playing parameters while playing.

3.2 Playing Parameters

This case study only analysed two playing parameters: blowing pressure and DC reed position. Previous studies [9, 11, 12] have shown the influence of blowing pressure, bite force and bite position on fundamental, sound level, and transient behaviour. Here, the DC reed position, which is related to both the player's bite on the reed and the blowing pressure, was used as a simple parameter for initial exploration.

Figure 6 is the boxplot calculated from the blowing pressure (the DC pressure in the mouth) of all the notes for the three EGs. The box shows the median for all notes and the variation is the note-to-note variation; variation within notes is not shown here. *Happy* shows the highest median blowing pressure (4.2 kPa) and largest variation in blowing pressure. The median blowing pressure for *deadpan* and *sad* are 3.6 kPa and 3.4 kPa, respectively, but *sad* shows a larger variation than *deadpan*. This is consistent with the observations from Figure 3, because the sound level of the radiated sound is expected to correlate with blowing pressure over the relatively low values used here [9]. The correlation coefficient is 0.56.

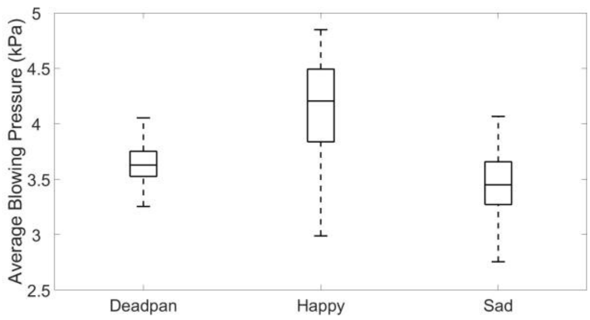


Figure 6. Boxplot of blowing pressure of all the notes for the three expressive goals: *happy*, *sad* and *deadpan*.

Figure 7 is the boxplot of DC reed displacements of all the notes for the three EGs. In Figure 7, zero on the y axis (not shown) corresponds to the reed position where the reed is at rest position without any displacement, and a more negative value corresponds to more reduced opening of the reed-mouthpiece aperture. (The player's bite produces a negative displacement.) *Happy* shows the least reduced aperture (lowest reed displacement, −0.89 mm) and largest variation in reed position among the three EGs, followed by *deadpan* (−0.92 mm) and *sad* (the most reduced opening, with displacement −0.94 mm). On its own, a higher blowing pressure tends to close the reed, giving a more negative reed displacement, so one might naïvely expect a more negative displacement for *happy*, where higher blowing pressure is used. However, the opposite is shown here. The explanation is that, in order to put the reed in an operating position to produce a sound with the similar sound level at lower blowing pressure, more lip force is required [9]. So, lip force needs to be increased further when playing with lower blowing pressure at lower sound level. This indicates that the clarinettist has applied more bite force when using lower blowing pressure to play *sad*.

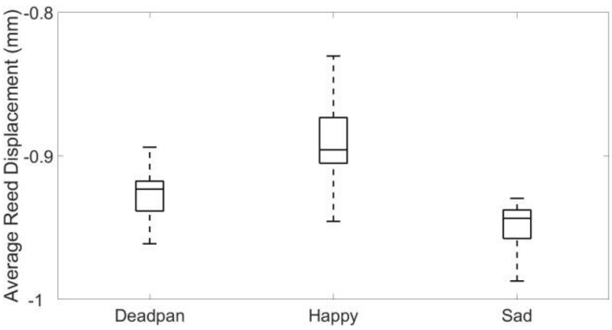


Figure 7. Boxplot of reed displacements of all the notes for the three expressive goals: *happy*, *sad* and *deadpan*. Higher (less negative) position means larger reed-mouthpiece aperture.

3.3 Inferential Statistical Analysis

	ANOVA	*happy* wrt *deadpan*	*sad* wrt *deadpan*	*sad* wrt *happy*
Sound level	F(2, 146) = 25.57 p < .001	↑	↓	↓
Tempo	F(2, 146) = 8.33 p < .001	ns	↓	↓
Spectral centroid	F(2, 146) = 9.67 p < .001	ns	↑	ns
Blowing pressure	F(2, 146) = 60.46 p < .001	↑	↓	↓
Reed position	F(2, 146) = 104.78 p < .001	↑	↓	↓

Table 1. Comparing five parameters of the three EGs using analysis of variance and post-hoc Tukey's honest significance test. ‘ns’ stands for non-significant result with adjusted p value larger than 0.05; ↑ and ↓ stand for larger and smaller, respectively. ‘wrt’ means ‘with respect to’.

Analysis of variance (ANOVA) and post-hoc Tukey's honest significance test were conducted to check further the significance of differences in the musical and playing parameters per note when comparing the three EGs. As shown in Table 1, most of the items are significantly different (with adjusted p value less than .05), except tempo and spectral centroid between *happy* and *deadpan*, and spectral centroid between *happy* and *sad*. These results statistically confirm the observations discussed earlier.

4. CONCLUSIONS

This preliminary case study involved one clarinettist participant and used the MIPCAT toolbox to investigate experimentally (1) the differences in the musical parameters of three EGs *happy*, *sad* and *deadpan* and (2) how the player conveys different EGs controlling the playing parameters, when performing the same musical excerpt ‘Happy Birthday to You’. Overall, notes played in the *happy* EG condition have high sound level and large standard deviation in sound level both within notes and across notes, fast tempo with medium variation, and moderate spectral change. This player used high blowing pressure with greatest variation and bite when interpreting *happy*. For the *deadpan* EG, the player selected medium blowing pressure and bite with little variation, resulting in a performance with medium sound level with minimum variation, fast and stable tempo, and minimum spectral change. For *sad*, low blowing pressure and bite with large variation were used to produce notes showing low sound level but large variation both within notes and across notes, slow tempo with large variation and large spectral variation. The analysis along the note sequence also revealed that *sad* has a very different pattern from *happy* and *deadpan*: the last two phrases have greater variation in the sound level and the peak sound level falling on the third last note (G4) instead of the highest note in the third phrase. Musical and playing features like these could be useful hints for music students to learn playing the music more expressively, especially when a larger sample of players has been measured.

While the present study focussed on just two of the measured playing parameters, MIPCAT offers potential in understanding how playing parameters interact to produce particular musical parameters. This will help to further bridge the gap between what the player is physically doing to the instrument during playing, and the aesthetic, musical product.

This study was limited to one player performing one musical excerpt and only the basic musical and playing parameters were investigated qualitatively. Thus the observations on comparing the three EGs could be unique for this particular player, e.g. similar tempo used for *deadpan* and *happy*. A future study with more players and detailed analysis would be ideal for identifying general regularities.

As a musical instrument performance capture and analysis toolbox, MIPCAT includes both hardware and software. The software part can process and analyse recordings of player performances captured by the hardware semi-automatically, reducing the work involved in the analyses reported here; it makes possible a more ambitious study on a group of clarinettist participants; this is currently being undertaken. Much of the software could, with relatively little modification, be applied to other instruments. For this reason, the software components are publicly available [15].

Acknowledgments
We warmly thank the Australian Research Council for support of this project, Yamaha for the clarinet, Légère for the reeds and the participant for the study.

5. REFERENCES

[1] A. Gabrielsson and P. N. Juslin, "Emotional expression in music performance: Between the performer's intention and the listener's experience," in Psychology of Music, vol. 24, no. 1, pp. 68-91, 1996.

[2] P. N. Juslin, "Cue utilization in communication of emotion in music performance: Relating performance to perception," in Journal of Experimental Psychology: Human Perception Performance, vol. 26, no. 6, p. 1797, 2000.

[3] P. N. Juslin, "Perceived emotional expression in synthesized performances of a short melody: Capturing the listener's judgment policy," in Musicae Scientiae, vol. 1, no. 2, pp. 225-256, 1997.

[4] P. N. Juslin, "Can results from studies of perceived expression in musical performances be generalized across response formats?" in Psychomusicology: A Journal of Research in Music Cognition, vol. 16, no. 1-2, p. 77, 1997.

[5] A. Gabrielsson and E. Lindström, "The role of structure in the musical expression of emotions," in Handbook of Music and Emotion: Theory, Research, Applications, vol. 367400, 2010, pp. 367-44.

[6] G. Kreutz, E. Schubert, and L. A. Mitchell, "Cognitive styles of music listening," in Music Perception, vol. 26, no. 1, pp. 57-73, 2008.

[7] G. P. Scavone, A. Lefebvre, and A. R. da Silva, "Measurement of vocal-tract influence during saxophone performance," in The Journal of the Acoustical Society of America, vol. 123, no. 4, pp. 2391-2400, 2008.

[8] P. Guillemain, C. Vergez, D. Ferrand, and A. Farcy, "An instrumented saxophone mouthpiece and its use to understand how an experienced musician plays," in Acta Acustica united with Acustica, vol. 96, no. 4, pp. 622-634, 2010.

[9] A. Almeida, D. George, J. Smith, and J. Wolfe, "The clarinet: How blowing pressure, lip force, lip position and reed “hardness” affect pitch, sound level, and spectrum," in The Journal of the Acoustical Society of America, vol. 134, no. 3, pp. 2247-2255, 2013.

[10] V. Chatziioannou and A. Hofmann, "Physics-based analysis of articulatory player actions in single-reed woodwind instruments," in Acta Acustica united with Acustica, vol. 101, no. 2, pp. 292-299, 2015.

[11] W. Li, A. Almeida, J. Smith, and J. Wolfe, "The effect of blowing pressure, lip force and tonguing on transients: A study using a clarinet-playing machine," in The Journal of the Acoustical Society of America, vol. 140, no. 2, pp. 1089-1100, 2016.

[12] W. Li, A. Almeida, J. Smith, and J. Wolfe, "How clarinettists articulate: The effect of blowing pressure and tonguing on initial and final transients," in The Journal of the Acoustical Society of America, vol. 139, no. 2, pp. 825-838, 2016.

[13] M. Pàmies-Vilà, A. Hofmann, and V. Chatziioannou, "Analysis of tonguing and blowing actions during clarinet performance," in Frontiers in Psychology, vol. 9, p. 617, 2018.

[14] A. Almeida, W. Li, E. Schubert, J. Smith, and J. Wolfe, "Expressive goals for performing musicians: The case of clarinetists," in Musicae Scientiae, p. 10298649221122155, 2022.

[15] A. Almeida, W. Li, E. Schubert, J. Smith, and J. Wolfe, "Recording and analysing physical control variables used in clarinet playing: A Musical Instrument Performance Capture and Analysis Toolbox (MIPCAT)," in Frontiers in Signal Processing, vol. 3, p. 1, 2023.

[16] A. Almedia, W. Li, E. Schubert, J. Wolfe, and J. Smith, "MIPCAT — A Music Instrument Performance Capture and Analysis Toolbox," in Stockholm Music Acoustics Conference, Stockholm, 2023.

[17] P. N. Juslin, "Emotional communication in music performance: A functionalist perspective and some data," in Music perception, 14(4), 383-418, 1997.

[18] E. Schubert and J. Wolfe, "Does timbral brightness scale with frequency and spectral centroid?" in Acta Acustica united with Acustica, vol. 92, no. 5, pp. 820-825, 2006.

[19] Y. Xu, A. Kelly, and C. Smillie, "Emotional expressions as communicative signals," in Prosody and Iconicity, 33–60, 2013.

Development of Vocal Vibrato Measures During Conservatory Training - A Longitudinal Retrospective Study

Reuben S. Walker[1,2], Mario Fleischer[1], Marie Bieber[2], and Dirk Mürbe[1,2]
[1]Department of Audiology and Phoniatrics,
Charité – Universitätsmedizin Berlin,
Corporate Member of Freie Universität Berlin
and Humboldt-Universität zu Berlin,
[2]Hochschule für Musik Carl Maria von Weber
Dresden, Germany
reuben-scott.walker@charite.de

ABSTRACT

Vibrato is a central component of western classical singing technique. As such, it is important for singers to have an understanding of the acoustic characteristics of vibrato and how those features could develop over a course of study.

Vibrato rate and extent were calculated automatically from recordings of sustained tones for 141 voice majors from the Hochschule für Musik Carl Maria von Weber Dresden, collected at the beginning and conclusion of university studies over a period from 2002 to 2021.

Median vibrato extent for all students increased from 19 cents, interquartile range (IQR) (12, 33) to 32 cents, IQR (20, 51). The median vibrato rate increased from 5.44 Hz, IQR (5.01, 5.78) to 5.59 Hz, IQR (5.08, 6.09) while the standard deviation of vibrato rate decreased from the initial recording to the final recording at graduation: 0.20 Hz IQR (0.09, 0.58) to 0.11 Hz, IQR (0.07, 0.28).

The classical voice majors considered in this study exhibited increased vibrato extent and vibrato rate stability during sustained tones at the conclusion of conservatory studies.

1. INTRODUCTION

Vibrato is a central feature of western classical singing education. Some singing styles are characterized by persistent vibrato and some by its absence, like multi-part ensemble singing [1] or certain jazz styles [2]. Musical theater and singers of contemporary commercial styles often begin by sustaining a note without perceptible vibrato for an extended period of time, then allowing the pitch to oscillate before moving to the next word or tone [3]. The listener perceives these oscillations as a combination of sound intensity and frequency fluctuations [4]. As sustaining a constant frequency is not possible, random frequency fluctuations happen naturally in both singing and speech [5]. Performing a "non-vibrato" sung tone requires training; singers specializing in styles with low vibrato extent in the literature had extensive vocal experience [3, 6].

Periodic frequency variations between 6 and 12 Hz are ever-present in speech [6]. These vocal tremors often become noticeable through an increased frequency extent in pathological cases, such as Parkinson's patients [7] or the aging voice [8]. Heightened emotional stress and anxiety have been linked to vocal frequency variation with conflicting results [9].

Perceptually, vibrato rate, vibrato extent, and cycle-to-cycle variation in vibrato rate interact as listeners determine which sung tones exhibit vibrato or are "non-vibrato [4]." Vibrato rate for commercial performances has shifted historically [10] but has been repeatedly measured between 5 Hz and 7 Hz among professional singers [2, 11]. Vibrato extent has been reported for mean-to-peak amplitude as approximately one semitone in professional singers [10, 12] and 110 cents in a theoretical model [13]. Vibrato extent was measured for peak-to-peak amplitude between 50 and 200 cents among professional opera singers [2].

From a technical perspective, vibrato is theorized to incorporate the development of muscle coordination between the thyroarytenoid and cricothyroid muscles [13]. Fluctuations in sound pressure level could be influenced by varying subglottal pressure [10] as well as by formant interaction [14]. Children's thyroarytenoid muscles are still developing [15] and exhibit more cycle-to-cycle frequency variation than adults [16]. Teaching maturing students to develop vibrato in their singing consists of encouraging them to experiment with allowing natural pitch fluctuations while incorporating exercises to develop the muscular antagonism necessary to increase the vibrato extent [14].

In the existing vibrato literature, students and pedagogues can find normative data from recordings of professional singers [3, 11], normative student data from larger samples [17], and small sample longitudinal data for vibrato development over a course of study [18]. However, there is a lack of longitudinal data describing vibrato development in a large sample of students. To address this, student recordings were analyzed from the Hochschule für Musik Carl Maria von Weber Dresden in order to provide a descriptive overview of student development with respect to vibrato rate, vibrato extent, and presence of vibrato in sustained tones.

2. METHODOLOGY

141 students (F = 82, M = 59) at the Hochschule für Musik Carl Maria von Weber Dresden were analyzed at the beginning and end of their studies. After being accepted to the conservatory, students were evaluated by a pedagogue with various quantitative and qualitative measures. Information gathered for this study included year of enrollment, course of study (classical or jazz/pop), gender, vocal register, and a subjective rating of vibrato stability. Only classical voice majors were included in the analysis.

Each student was recorded at regular intervals during their studies. As the recordings were made at different times during the winter and summer semesters, singers were only selected for analysis who had final recording dates at least 2.5 calendar years removed from their initial recordings to approximate the three year minimum duration of the bachelors degree. The median duration from first to final recording was 4.4 years. The singers were recorded performing seven different vocal exercises. For this study, an ascending and descending major triad with a sustained fifth was selected. The exercise was sung on an /aː/ vowel. High voices performed the exercise in A major while medium/low voices performed the exercise in G major.

The pitch contour of the exercise (see Figure 1) was calculated using the PRAAT Parselmouth [19] library in python and the middle fifty percent of the highest sustained tone was isolated. The mean and standard deviation of vibrato rate were calculated by performing an autocorrelation over rolling windows with a one second duration. If the extrema of the autocorrelation function exceeded the 95% confidence intervals between the equivalent lags for 3 Hz and 10 Hz, the vibrato rate was recorded for the maximum lag and the mean vibrato extent was recorded for the window. If the extrema did not exceed the confidence intervals, the vibrato rate and extent were recorded as null values. The vibrato extent was calculated by taking the mean of pitch contour prominences separated by a minimum distance (75% of the corresponding wavelength for that window). In order to avoid partial amplitudes, the initial and final peak of each window were excluded. The peak amplitude was converted to cents with reference to the mean frequency. Mean and standard deviation values were then aggregated from the series of rolling windows for both vibrato rate and vibrato extent. Finally, a vibrato percentage was calculated from all portions of the sustained tone that had recorded non-null values. If a vibrato rate was measurable for all rolling windows of the pitch contour, the student would have a vibrato percentage of 100%. Sample code is provided on GitHub [20].

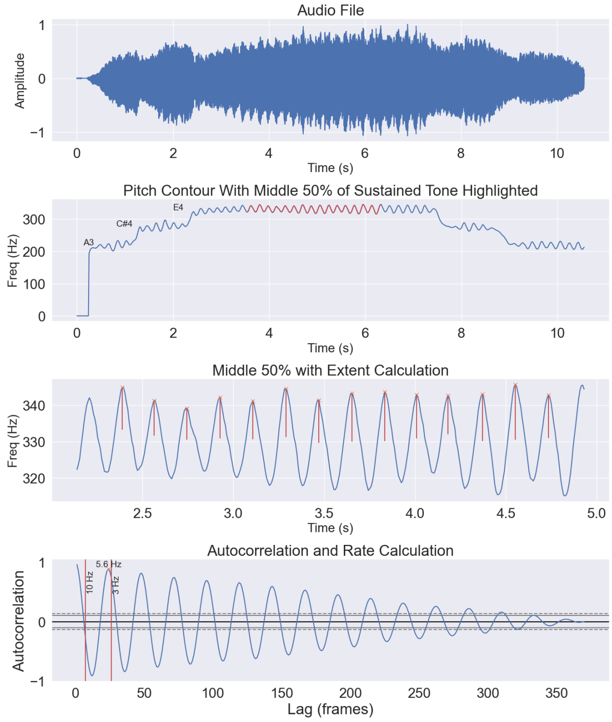


Figure 1. From the initial audio file, the middle 50% of the highest sustained tone was isolated from the pitch contour, calculated with the PRAAT Parselmouth library. The vibrato rate was calculated by finding the peak in the autocorrelation function between 3 Hz and 10 Hz. The vibrato extent was averaged from the peak prominences of the pitch contour and converted to cents with reference to the mean pitch.

3. RESULTS

Here we provide a descriptive overview of vibrato rate, vibrato extent, and presence of vibrato in sung tones (vibrato percentage) at the beginning and end of studies. The measures are presented together in Table 1.

3.1 Normality Tests

As all of the measured values were bounded, a Shapiro-Wilkes test was used to check the distributions for normality before determining the appropriate descriptive analysis. With the exception of vibrato rate at graduation ($W = 0.99$, $p = 0.29$), the remaining measures all deviated significantly from normality ($p < 0.05$). As such, we report the median values and inter-quartile range (IQR), presented as the 25% and 75% values of the distribution.

3.2 Vibrato rate

The group median vibrato rate at the beginning of studies was 5.44 Hz, with 50% of the measured values (IQR) falling between 5.01 Hz and 5.78 Hz. At graduation the median vibrato rate was 5.59 Hz, IQR (5.08, 6.09) (see Figure 2A). At the beginning of studies, the median intrasubject standard deviation of vibrato rate was 0.20 Hz, IQR (0.09, 0.58). At graduation the median value was 0.11

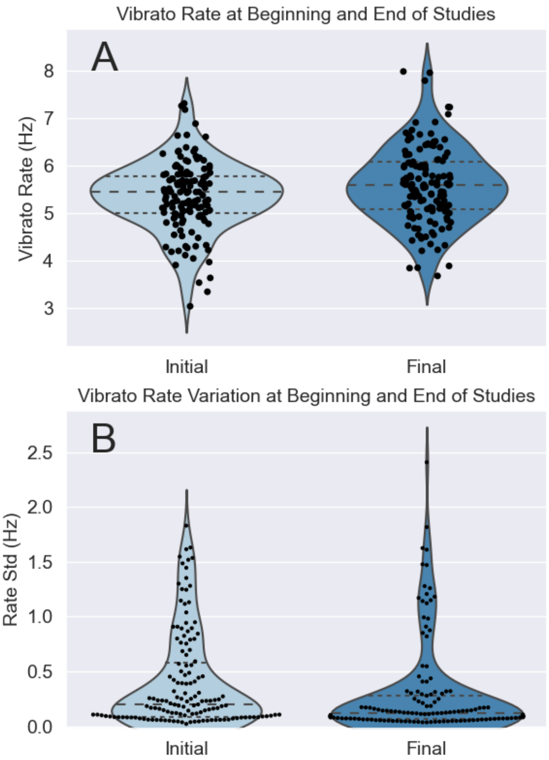


Figure 2. **(A)** The group median vibrato rate increased from 5.44 Hz, IQR (5.01, 5.78) at the beginning of studies to 5.59 Hz, IQR (5.08, 6.09) at graduation. **(B)** Median intrasubject standard deviation of vibrato rate decreased from the initial recording to the final recording at graduation 0.20 Hz, IQR (0.09, 0.58) to 0.11 Hz, IQ (0.07, 0.28).

Hz, IQR (0.07, 0.28). Vibrato rate variance is presented in Figure 2B.

3.3 Vibrato Extent

Vibrato extent was calculated as mean-to-peak displacement of the pitch contour. As seen in Figure 3A, median vibrato extent at the beginning of studies was 19 cents, IQR (12, 33). At graduation, the median vibrato extent was 32 cents, IQR (20, 51). Median intrasubject standard deviation of vibrato extent at the beginning of studies was 2.9 cents IQR (1.9, 4.1). At graduation, the median value was 3.6 cents IQR (2.3, 5.5). Vibrato extent variation is presented in Figure 3B.

3.4 Vibrato Percentage

At the beginning of studies, the median vibrato percentage for the sample was 100%, IQR (94, 100), with 97 of 141 subjects recording 100% vibrato (see Figure 4). At graduation, 116 of 141 singers recorded vibrato measures for the entirety of the sustained tone, Mdn = 100%, IQR (100, 100).

The development of vibrato rate, vibrato extent, and vibrato percentage is visible in Figure 5. The vibrato characteristics of singers whose vibrato ratings were “unstable” or “no vibrato” at the beginning of studies more closely

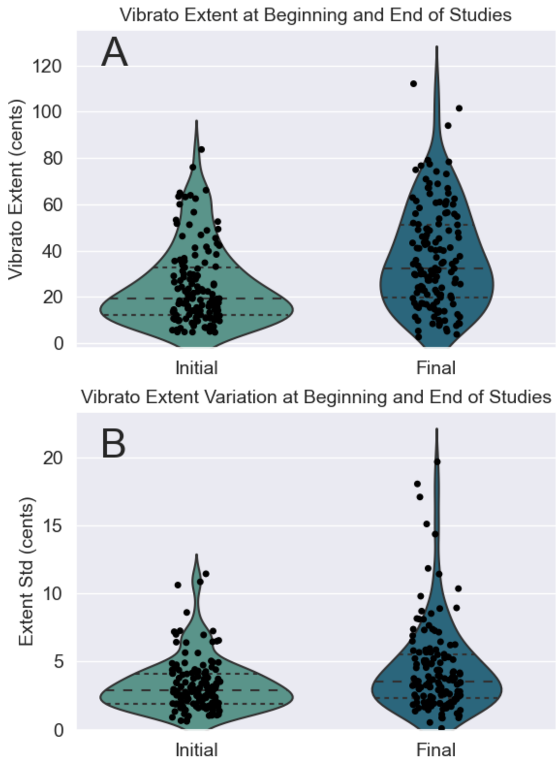


Figure 3. **(A)** Median vibrato extent for all students increased over the course of study from 19 cents, IQR (12, 33) to 32 cents, IQR (20, 51). **(B)**: Median intrasubject standard deviation of vibrato extent increased from the initial recording to the final recording at graduation 2.9 cents IQR (1.9, 4.1) to 3.6 cents IQR (2.3, 5.5).

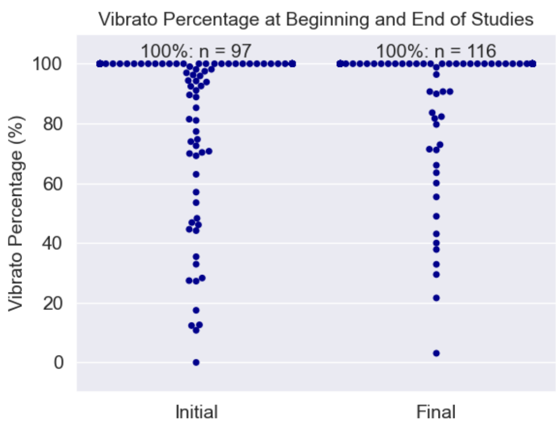


Figure 4. At the beginning of studies, the median vibrato percentage for the sample was 100% IQR (94, 100), with 97 of 141 subjects recording 100% vibrato. At graduation, the number of singers who registered 100% vibrato for the sample had increased to 116, Mdn = 100%, IQR (100, 100).

approximated the “stable” vibrato group at graduation.

Table 1. Vibrato measures at the beginning of studies and at graduation.

Vibrato Measure	Initial Median	(IQR)	Graduation Median	(IQR)
Rate (Hz.)	5.44	(5.01, 5.78)	5.59	(5.08, 6.09)
Intrasubj. SD (Hz)	0.20	(0.09, 0.58)	0.11	(0.07, 0.28)
Extent (cents)	19	(12, 33)	32	(20, 51)
Intrasubj. SD (cents)	2.9	(1.9, 4.1)	3.6	(2.3, 5.5)
Percentage (%)	100	(94, 100)	100	(100,100)

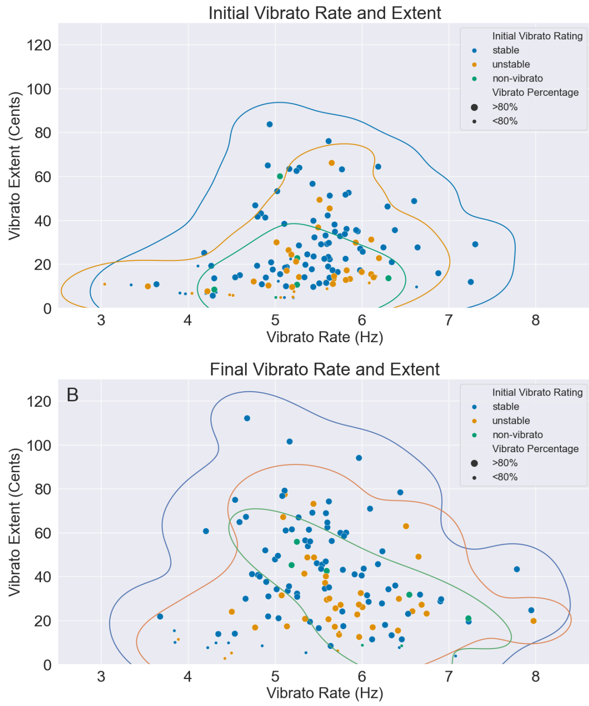


Figure 5. Vibrato extent plotted against frequency for initial **(A)** and final **(B)** recordings. Contours represent a probability distribution (at the 0.05 level) for groupings provided by the vibrato ratings at the beginning of studies (stable/unstable/non-vibrato). Larger dots represent students whose vibrato rate was recognizable for more than 80% of the sustained sample. Smaller dots represent students whose vibrato rate was recognizable for less than 80% of the sustained sample. By graduation, the contour of the “unstable” and “non-vibrato” singers better approximates that of the “stable” singers.

3.5 Singers with Fast/Slow Vibrato

As shown in Figure 6, the 54 singers with initial vibrato rate below 5.2 Hz increased their vibrato rate by 0.6 Hz, 95% CI [0.4, 0.9]. The 35 singers with initial vibrato rate above 5.8 Hz did not decrease their vibrato rate beyond the margin of error, 95% CI[-0.4, 0.2] Hz.

4. DISCUSSION

Although the mean vibrato rate was greater at graduation than at the beginning of studies, many of the low frequency values were recorded in student samples initially rated as “unstable” or “no vibrato” and exhibited such low values for vibrato percent and vibrato extent that the extracted frequency fluctuations that were recorded as “vibrato rate” should be considered with caution. Among the 69 students exhibiting more than 20 cents of vibrato extent both at the beginning of studies and at graduation, the difference in vibrato rate did not change (5.5 Hz 95% CI [5.42, 5.66] to 5.5 Hz 95% CI [5.34, 5.67]). Future studies should consider a threshold for both extent and rate of pitch fluctuations before categorizing pitch fluctuations as “vibrato”.

The observed increase in vibrato extent was mainly driven by students with low vibrato extent entering into the range of perceivable extent, although there were singers whose mean vibrato extent at graduation exceeded 100 cents. Ideally, instructors would hope that a region in the vibrato extent and vibrato rate field would serve as an attractor with continued study. The low values of vibrato extent compared to those of professional operatic singers in the literature could be due to a number of factors. Firstly, the stylistic specialization of the students was not available for the analysis. Singers specializing in baroque concert repertoire or contemporary choral singing would not exhibit the same vibrato extent as singers specializing in romantic operatic repertoire. Secondly, although the chosen vocal exercise provided a reasonably controlled approximation of

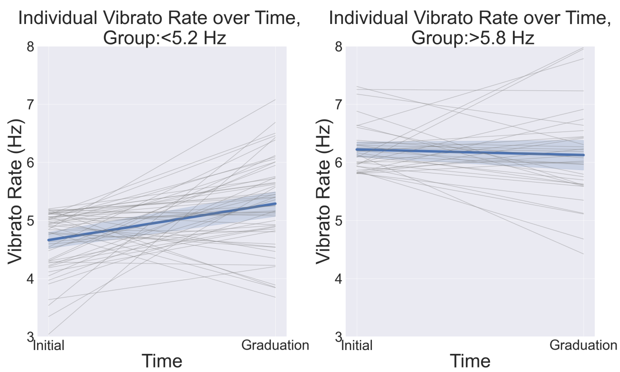


Figure 6. Individual vibrato rate over time for slow and fast subgroups. The blue line represents the mean difference while the shaded region represents the 95% confidence interval. The slow subgroup (n=54) increased their vibrato rate by 0.6 Hz, 95% CI [0.4, 0.9]. The fast subgroup (n=35) did not decrease their vibrato rate beyond the margin of error, 95% CI[-0.4, 0.2] Hz.

sustained singing in a comfortable range, it was still an approximation. Students were not explicitly instructed to sing with or without vibrato and may have limited their vibrato extent in order to increase pitch accuracy in an unfamiliar (though straightforward) exercise.

The increase in vibrato extent variability was counterintuitive. As standard deviation in vibrato extent was an intra-subject measure computed via a rolling window, it is possible that the window chosen for sustained phonation (middle 50% of the sustained tone) captured stylistic features, such as an increase in vibrato extent toward the end of the sustained tone. Additionally, four of the five singers with the highest increases in vibrato extent standard deviation had sustained tones at the beginning of studies with less than twenty cents of vibrato extent, a value more likely to be classified as "non-vibrato" by listeners [4]. Those singers' vibrato extent increased to between 30 and 77 cents by the end of studies. An increase in vibrato extent variability could be normal for a population of conservatory singers learning to sing with consistent classical vibrato.

Though the increase in vibrato rate for the "slow vibrato" subgroup that began with a frequency fluctuations less than 5.2 Hz was encouraging, it is again important to note that some of those singers began with values taken from just a small portion of their sustained phonation with low values for the vibrato extent. The lack of regression to the mean was surprising for the group of singers that began with vibrato greater than 5.8 Hz. In this "fast vibrato" subset, there were three singers whose vibrato increased to more than 7.7 Hz over the course of study.

5. CONCLUSION

Although vibrato is often an ornamental consideration in vocal instruction, non-normative extent and frequency can be disqualifying factors in an audition setting and often serve as a first sign of vocal decline. Over the course of conservatory training at the Hochschule für Musik Carl Maria von Weber Dresden, vocal development was characterized by increases in vibrato extent and decreases in variation of the vibrato rate. The range of normative values and trends should serve as guidelines for students and teachers evaluating potential outcomes during training.

Acknowledgments

Thanks go out to Svante Granqvist, who provided software to convert the audio files for analysis and to Paul Boersma and David Weenik for the use of the PRAAT Parselmouth library. We thank Prof. Hartmut Zabel for discussions on the intake and recording procedure at the Hochschule für Musik Carl Maria von Weber Dresden.

6. REFERENCES

[1] J. Dai and S. Dixon, "Intonation trajectories within tones in unaccompanied soprano, alto, tenor, bass quartet singing," *The Journal of the Acoustical Society of America*, vol. 146, no. 2, pp. 1005–1014, 2019.

[2] C. Manfredi, D. Barbagallo, G. Baracca, S. Orlandi, A. Bandini, and P. H. Dejonckere, "Automatic assessment of acoustic parameters of the singing voice: application to professional western operatic and jazz singers," *Journal of Voice*, vol. 29, no. 4, pp. 517–e1, 2015.

[3] A. S. Becker and P. J. Watson, "The use of vibrato in belt and legit styles of singing in professional female musical-theater performers," *Journal of Voice*, 2022.

[4] R. Wooding and J. Nix, "Perception of non-vibrato sung tones: a pilot study," *Journal of Voice*, vol. 30, no. 6, pp. 762–e15, 2016.

[5] J. Verstraete, G. Forrez, P. Mertens, and F. Debruyne, "The effect of sustained phonation at high and low pitch on vocal jitter and shimmer," *Folia Phoniatrica et Logopaedica*, vol. 45, no. 5, pp. 223–228, 1993.

[6] C. F. Sinclair, L. E. Gurey, and A. Blitzer, "11 neurologic and neuromuscular diseases of the larynx," *Clinical Laryngology*, p. 97, 2014.

[7] L. A. Ramig and T. Shipp, "Comparative measures of vocal tremor and vocal vibrato," *Journal of Voice*, vol. 1, no. 2, pp. 162–167, 1987.

[8] M. Brückl, *Altersbedingte Veränderungen der Stimme und Sprechweise von Frauen*. Logos Verlag Berlin GmbH, 2011, vol. 7.

[9] C. L. Giddens, K. W. Barron, J. Byrd-Craven, K. F. Clark, and A. S. Winter, "Vocal indices of stress: a review," *Journal of voice*, vol. 27, no. 3, pp. 390–e21, 2013.

[10] J. Sundberg, "Acoustic and psychoacoustic aspects of vocal vibrato," *Vibrato*, pp. 35–62, 1995.

[11] E. Prame, "Measurements of the vibrato rate of ten singers," *The journal of the Acoustical Society of America*, vol. 96, no. 4, pp. 1979–1984, 1994.

[12] J. Hakes, T. Shipp, and E. T. Doherty, "Acoustic characteristics of vocal oscillations: vibrato, exaggerated vibrato, trill, and trillo," *Journal of Voice*, vol. 1, no. 4, pp. 326–331, 1988.

[13] I. R. Titze, B. Story, M. Smith, and R. Long, "A reflex resonance model of vocal vibrato," *The Journal of the Acoustical Society of America*, vol. 111, no. 5, pp. 2272–2282, 2002.

[14] J. Nix, "Shaken, not stirred: Practical ideas for addressing vibrato and nonvibrato singing in the studio and the choral rehearsal," *Journal of Singing*, vol. 70, no. 4, pp. 411–418, 2014.

[15] M. Hirano, "Growth, development, and aging of human vocal folds," *Vocal fold physiology*, pp. 22–43, 1983.

[16] M. Döllinger, D. Dubrovskiy, and R. Patel, "Spatiotemporal analysis of vocal fold vibrations between children and adults," *The Laryngoscope*, vol. 122, no. 11, pp. 2511–2518, 2012.

[17] J. Nix, N. Perna, K. James, and S. Allen, "Vibrato rate and extent in college music majors: a multicenter study," *Journal of Voice*, vol. 30, no. 6, pp. 756–e31, 2016.

[18] D. Mürbe, T. Zahnert, E. Kuhlisch, and J. Sundberg, "Effects of professional singing education on vocal vibrato—a longitudinal study," *Journal of Voice*, vol. 21, no. 6, pp. 683–688, 2007.

[19] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]," Version 6.1.38, retrieved 2 January 2021 http://www.praat.org/, 2021.

[20] R. S. Walker, "Vibrato frequency and extent calculation," 2023. [Online]. Available: https://github.com/reubenwalker/VibratoStudy.git

IMPACT OF SINGING TOGETHERNESS AND TASK COMPLEXITY ON CHORISTERS' BODY MOTION

Sara D'Amario
mdw – University of Music and Performing Arts Vienna
damario@mdw.ac.at

Sten Ternström
KTH Royal Institute of Technology
stern@kth.se

Werner Goebl
mdw – University of Music and Performing Arts Vienna
goebl@mdw.ac.at

Laura Bishop
University of Oslo
laura.bishop@imw.uio.no

ABSTRACT

We examined the impact of the perception of singing togetherness, as indexed by the spectral envelope of the sound, and task complexity on choristers' body motion, as they performed in duo with a pre-recorded tune presented over a loudspeaker. Fifteen experienced choral singers first manipulated the spectral filter settings of the tune in order to identify the recordings they felt *most* and *not at all* together with. Then, they sang the tune in unison and canon along with the recordings featuring the chosen filter settings. Audio and motion capture data of the musicians' upper bodies during repeated performances of the same tune were collected. Results demonstrate that wrist motion was more periodic, singer posture more open, and the overall quantity of body motion higher when singing in unison than in canon; singing togetherness did not impact body motion. The current findings suggest that some body movements may support choral performance, depending on the complexity of the task condition.

1. INTRODUCTION

Musicians' body motion is a fundamental aspect of music performance. Certain movements are intended to produce and manipulate the sounds; others are employed to communicate expressive ideas with the co-performer(s) and the audience [1, 2]. Studies on musicians' body motion have shown that the head motion of instrumental players can be related to the emotional intentions that musicians aim to convey [3]. Very few investigations have analysed singers' body movements, which are related to the music background of the musicians: interpersonal coordination of movements was found to be higher within an expert choir than in a novice choir [4].

A recent investigation has also shown a relationship between musicians' body motion and the level of togetherness in musical duos, judged by novices and semi-professional musicians [5]. Specifically, it was found that novices judged more together duo performances with higher than lower similarity in right arm motion, and semi-professional musicians judged more together duo performances with higher than lower chest motion coordination.

These results suggest that performers' perception of togetherness with a co-performer can also be linked to their body motion.

Togetherness is particularly relevant to choirs, where singers are often required to blend pitch, intensity, timbre and timing of their tones with those of the co-performers. Choristers anecdotally assert that the ease of singing together with a co-performer in a choir can vary greatly with the individual. This sensation might be related to the complementarity of the spectral components of the sound, since it has been found that singers in a barbershop quartet seem to spread out their formant frequencies rather than align them [6]. This demonstrates that singers tend to control their voice spectrum, probably, to blend their voice with the ensemble's and this might also affect their body motion. However, the effects of the perception of togetherness (related to the sound spectrum) on singers' body motion have not yet been investigated.

Furthermore, head movements are also related to the complexity of task production. It has been shown that side-to-side head movements were more variable when singing duos performed the same piece in Rounds (i.e., one singer entering in canon with an eight-tone offset) than in unison (i.e., singing simultaneously identical pitches) [7]. The greater difficulty of task production might also affect the motion of other body parts. Timing and pitch disruptions posited by a canon performance might induce larger overall body motion and lesser periodicity in wrist's movement than singing in unison.

The current study aims to investigate the chorister's upper body motion when singing in duo along to a pre-recorded tune presented over the loudspeaker. Specifically, we tested the hypotheses that the perception of togetherness with a co-performer and the complexity of the task production affects the chorister's body motion during singing duo performances.

2. METHODS

2.1 Participants

Fifteen trained singers (age M=31.2 years old, SD=11.2 years; 5 females, 10 males) with advanced choral experience took part in the study. They had on average 9.5 years of formal training (SD=4.3 years) and practiced on average 1.5 hours per day (SD=0.7 hours). They all self-reported having normal hearing, and three of them having perfect pitch.

Ethical approval for the study (with reference EK Nr: 05/2020) was obtained from the Ethics Committee at mdw.

2.2 Method

The experiment was a within-subject design comprising two stages: stage A, preferences (focused on singers' perception of singing togetherness/match) and stage B, performance (centered on singers' performance based on togetherness preferences collected during stage A). Each stage featured a 2 (parts: part 1 and part 2) x 2 (singing modes: unison and canon) x 2 (togetherness: most and not at all together) x 2 (takes: take 1 and take 2, i.e. repetitions of the same condition) design, with a total of 16 trials per participant. The order of stage and part was fixed. The order of condition in stage A was fixed within part; first, two repetitions in unison then in canon of the best perceived match; then, two consecutive repetitions in unison then canon of the worst perceived match. The order of conditions in stage B was randomized within part. Results from stage A are out of the scope of the present study and will be reported elsewhere.

2.3 Experimental Set-up and Apparatus

The experiment was set in a large, multipurpose room at mdw, equipped with a 12-camera (Prime 13) OptiTrack motion capture (MoCap) system, suspended from the room ceiling.

Choristers stood at about the center of the body motion rig and at about 70 cm from a loudspeaker (Genelec model 8020C) placed at their right side. They were facing an imaginary audience whilst two researchers were sitting at their desks behind the singers to record audio and body motion data. For stage A, a screen was also placed at the front-left of the participants to display visual instructions, as well as a MIDI controller device (Native Instruments, model 4control). The latter one was equipped with two unmarked rotary knobs that the participants used to control the filter settings.

Choristers' upper body motion was tracked at 240 frames per second using the MoCap system. Each singer was fitted with 11 reflective markers: three on the head, one on the chest and back, and one for each shoulder, elbow and wrist. Three additional markers were placed over the loudspeaker (as shown in Figure 1).

A total of five microphones were used to collect the audio recordings: one 'backstage' microphone (Neumann KM A P48) to synchronize motion and audio recordings; one reference microphone (Behringer model ECM8000), mounted in front of the singer; a binaural microphone (DPA model 4560) with the capsules placed near the singer's ear canal; and, two head-mounted microphones (DPA model 4066), one placed on the singer's and one on the loudspeaker over the side next to the participant. Furthermore, electroglottography (EGG) electrodes were placed over the singer's neck to analyse intonation without crosstalk from the loudspeaker. The analysis of the audio recordings is out of the scope of this study and those results will be reported elsewhere.

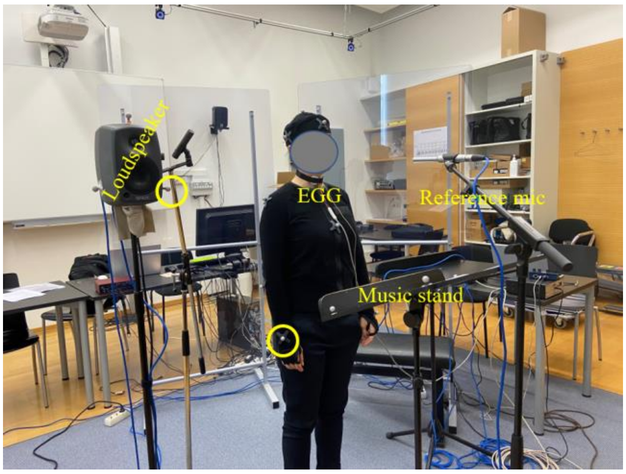


Figure 1. Example of the experimental set-up, showing some of the motion capture markers placed on the singer's right wrist and the loudspeaker (circled). The figure also displays the electroglottography (EGG) electrodes on the singer's neck as well as the reference microphone and the music stand placed in front of the singer.

Audio from the backstage microphone was routed into a multichannel audio interface (Focusrite Scarlett 18i8) and recorded using Ableton Live at 44.1 KHz and 32-bit depth using a desktop PC. The remaining audio and the MIDI data were routed to another digital interface (RME model UFX II) and recorded using a Microsoft Book 2 laptop.

2.4 Stimuli

A short tune (as shown in Figure 2) was composed for the study, so it was easy to be learned during a short practice session and could be sung in unison and also in canon with a two-bars offset.

The tune was pre-recorded by two professional singers (one male and one female) with extensive experience singing in choirs. They were instructed to perform as they would in a choir, with a limited vibrato. Four metronome beats at 90 BPM cued the tempo just before the beginning of the singing performance. The tune was pre-recorded so the participants taking part in the study could sing in duo along with it in unison and in canon.

2.5 Procedure

Participants were first given an oral and written explanation of the experiment, and gave written consent to participate in the study. Then, stage A was presented: singers listened to a set of 16 recordings whilst modifying the spectral components of the sound by rotating the two MIDI knobs, until they found the sound they felt was most/not at all together with, based on the researchers' instructions.

Two parametric filters were employed, centered on 2.7 kHz and 6.2 kHz, with an adjustable gain of ±15 dB. The frequency band at 2.7 kHz is related to the singer's formant and, when the level is very high in this region, the voice timbre is perceived as "piercing" or "projecting". In the frequency band at 6.2 kHz, voice timbre with a high level is perceived as "clear" "and "intimate" whilst a low level gives the impression that the sound is occluded. The filter settings of the chosen sounds were saved.

Figure 2. Piece composed for the study, highlighting the beginning note of each of its four phrases.

At the completion of stage A, stage B was presented and choristers sang the short tune in duo with the pre-recorded performance, outputted with the filter settings chosen in the previous stage. Pre-recorded stimuli were gender-matched with the participants.

To avoid fatigue, participants were given a 5 minutes rest between stages, and also between parts in stage B.

2.6 Analysis

To investigate the impact of the perception of togetherness and task complexity on choristers' body motion, first motion capture data were pre-processed: data of all markers were smoothed and velocity derived using a Savitzky-Golay filter (polynomial order 3, window size 25). The three-dimensional (3D) velocity and position data were also computed as the Euclidean norm of the smoothed velocity and position data (respectively). Then, the following metrics were computed:

- Quantity of motion (*QoM*), indicating the total energy of the singer during the performance
- Singer's posture, referring to the openness of the upper body
- Distance between the singer and the loudspeaker
- Periodicity of the singer's wrist and head motion

To compute *QoM*, 3D velocity curves were summed per second across markers for each singer and trial, then averaged across time stamps within phrases. Since the distributions of body motion velocities were right-skewed, *QoM* computation was based on the log of velocity data $[\log + 0.1]$.

Singers' posture was investigated by summing the 3D distance between the chest and the front head, and between the chest and all the peripheral joints under investigation (i.e., left and right shoulder and elbow). The distance between the singer and the loudspeaker was calculated as the standardized 3D distance from the singer's front head marker and that placed on the loudspeaker near the singer. These distances were first averaged per second, and then averaged within phrases.

The periodicity in the singers' front head and wrist motion was conceptualized in terms of power of the wavelet transforms (*WT*) of the 3D velocity curves of each marker of interest (i.e., front head and left and right wrist), as shown in the example of Figure 3. First *WT* power data were extracted in the range between 0.5 to 11 s using [8]; this broad band was set in line with the phrase structure and the tempo of the piece, and included periods ranging from 1 beat (mean duration = 0.667 s) to 4 bars (mean duration = 10.36 s). Then, within this broad band, the period with the strongest power was identified (for each marker, trial and singer) and the average *WT* power was calculated within a narrow band centered around this period with the strongest power with a width of $\pm$ 1 beat. Ultimately, *WT* data were averaged across timestamps within phrases.

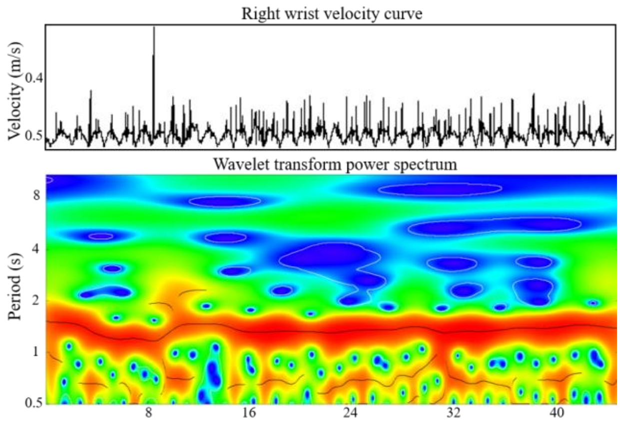


Figure 3. Example heat plot of the wavelet transform (*WT*) power spectrum (at the bottom) as computed from the right wrist 3D velocity curve (at the top) for a chorister performing the short tune in unison along with a pre-recorded performance of the same piece. The warmer the map, the more periodic the signal is.

Ultimately, mixed linear models were implemented to investigate the effects of the perception of togetherness, i.e. most and not at all together, and task complexity, i.e. singing in unison versus canon (the explanatory variables), on the above body motion metrics (the response variables). Participant and trial number were entered in the models as fully crossed random effects.

3. RESULTS

3.1 Perception of Togetherness

Perception of togetherness did not predict *QoM*, the singers' posture, or the distance between the singers' front head and the loudspeaker, as shown in Figure 4. In addition, perception of togetherness did not predict the periodicity of the right and left wrist motion, as shown in Figure 5, or the periodicity of the head motion.

3.2 Task Complexity: Singing in Unison and Canon

Task complexity predicted *QoM*, which was lower when singing in unison than when singing in canon ($\beta = -0.09, 95\%\ CI\ [-0.14, -0.05], t(449) = -4.13, p < 0.001$). It predicted also singers' posture, which was more open when singing in unison than in canon ($\beta = 0.04, 95\%\ CI\ [0.01, 0.06], t(234) = 2.95, p < 0.01$).

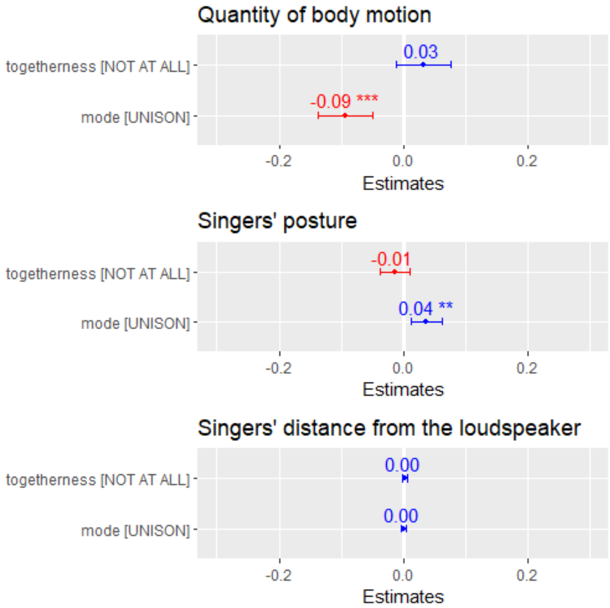


Figure 4. Estimates (on the x-axis) of togetherness perception and singing mode effects on singers' distance from the loudspeaker, posture and quantity of body motion (from the bottom up). Estimates are given above with reference to the specified base level of the factor (i.e., not at all *versus* most together, and unison *versus* canon). These estimates represent the difference in the predicted values of the response variable between the levels of the explanatory variable. For example, singers' quantity of motion (*QoM*) when performing in unison was on average 0.09 units lesser than when singing in canon. $*** \, p < 0.001$; $** \, p < 0.01$; $* \, p < 0.05$

Task complexity also predicted the periodicity of the left and right wrist motion, as shown in Figure 5. Left and right wrists were more periodic when singing in unison than when singing in canon ($\beta = 5.93, 95\%\ CI\ [1.15{,}10.7], t(234) = 2.45, p < 0.05$; and, $\beta = 6.67, 95\%\ CI\ [2.14{,}11.21], t(234) = 2.9, p < 0.01$; respectively). However, the periodicity of the head motion was not a significant predictor of unison and canon singing.

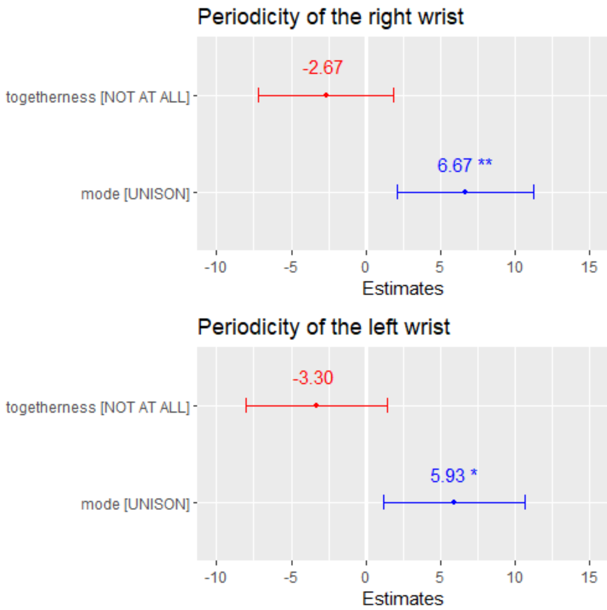


Figure 5. Estimates (on the x-axis) of togetherness perception and singing mode effects on the periodicity of the right and left wrist (from the bottom up). Estimates are given above with reference to the specified base level of the factor (i.e., not at all *versus* most together, and unison *versus* canon). $** \, p < 0.01$; $* \, p < 0.05$.

4. DISCUSSION AND CONCLUSIONS

This study examined choristers' body motion while performing in duo along with a pre-recorded co-performer. Specifically, it investigated the effects of togetherness sensation and task complexity on body motion.

We did not find evidence of a significant impact of togetherness perception, as indexed by the spectral envelope of the sound, on singers' body motion. This finding suggests that togetherness sensation might be irrelevant to body motion. However, it might also be that the singer-loudspeaker paradigm implemented in this study did not reflect the interpersonal dynamics and interactions typical of a small singing ensemble. Future ecologically valid investigations are needed to investigate the impact of togetherness perception on singers' body motion during ensemble singing.

The effects of task complexity on singers' body motion were manifested in a higher quantity of motion and lesser periodicity in wrist motion when singers performed in canon rather than in unison. These results extend previous investigations, showing higher variability in singing duos' head movements and slower tempi when singing the same piece at a time off-set [7]. Furthermore, these findings also expand studies on note-to-note synchronization of sound during piano duo performances demonstrating larger asynchronies when playing the same piece at a temporal offset rather than in unison [9]. Interestingly, during the participants' debriefing, they reported that singing in canon was more difficult than singing in unison, thus confirming differences between task performances as measured in the body motion.

Furthermore, singers' posture was more open when singing in unison and canon. Body posture in dyadic interactions can be related to interpersonal perception, and it has been shown that participants with an expanded body posture rated themselves higher on agency than those who displayed a restricted or neutral posture [10]. The fact that our participants exhibited a more expanded posture when singing in unison might reflect the need to stand more clearly during unison performances so their voices can be heard more easily.

The current study focused on the univariate analysis of dyadic interactions, i.e. how a singer reacts when performing along to a pre-recorded tune presented over a loudspeaker. This was done to control for the effects of the spectral features of the sound of one singer on the co-performer's body motion. Singer's body motion might exhibit more variability when performing with a real singer, due to continuous adaptations. Future investigations might shed more light in this respect, by analyzing the bivariate aspects of interpersonal coordination in choirs.

In summary, this study has shown that choristers' body motion, posture and distance from the co-performer can be impacted by the perception of togetherness with a co-performer and the complexity of the singing performance, suggesting that certain body motions may facilitate choral performance depending on the musical context. These results are relevant to choral pedagogy aiming at refining rehearsal techniques that support choral performance excellence.

Acknowledgments

We thank Christian Herbst, Josipa Bainac and Judith Kopecky for generously lending the electroglottograph, digital audio interface and microphones during data collection. This research is supported by the Austrian Science Fund, project P32642, and the University of Oslo and the Research Council of Norway through its Centres of Excellence scheme, project number 262762. Author S.T. participated on KTH faculty funds.

REFERENCES

[1] A. R. Jensenius, M. M. Wanderley, R. Goday, and M. Leman, "Musical gestures: concepts and methods in research," in *Musical Gestures: Sound, Movement, and Meaning*, R. Goday and M. Leman, Eds., NewYork, NY: Routledge, pp. 12–35, 2010.

[2] J. Davidson, J. S. Correira, "Body movements" in *The science and psychology of music performance: Creative Strategies for teaching and learning*, R. Parncutt, and G. McPherson, Eds., Oxford: Oxford University Press, pp. 237–250, 2002.

[3] A. Chang, H. E. Kragness, S. R. Livingstone, D. J. Bosnyak, and L. J. Trainor, "Body sway reflects joint emotional expression in music ensemble performance," in Scientific Reports, 9: 205, 2019. doi: 10.1038/s41598-018-36358-4

[4] T. Himberg, and M. Thompson, "Group synchronization of coordinated movements in a cross-cultural choir workshop," in Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM), pp. 175-180, 2009.

[5] S. D'Amario, W. Goebl, and L. Bishop, "Judgment of togetherness in performances by musical duos," in Frontiers in Psychology, 13: 997752, 2022.

[6] G Kalin (2005). "Formant frequency adjustment in barbershop quartet singing," M.Sc. thesis, Dept of Speech, Music and Hearing, KTH Royal Institute of Technology, 2005. Available online at http://www.speech.kth.se/prod/publications/files/1684.pdf .

[7] C. Palmer, F. Spidle, E. Koopmans, and P. Schubert, "Ears, heads, and eyes: When singers synchronise," Quarterly Journal of Experimental Psychology, vol. 72, 2019, no. 9, pp. 2272–2287, 2019.

[8] A. Roesch and H. Schmidbauer, "Waveletcomp: Computational Wavelet", 2018. Available online at: https://CRAN.R-project.org/ package=WaveletComp

[9] A. Zamm, P. Q. Pfordresher, and C. Palmer, "Temporal Coordination in joint music performance: Effects of endogenous rhythms and auditory feedback," Experimental Brain Research, vol. 233, no. 2, pp. 607–615, 2014.

[10] A. E. Abele and V. Yzerbyt, "Body posture and interpersonal perception in a dyadic interaction: A big two analysis," European Journal of Social Psychology, vol. 51, no. 1, pp. 23–39, 2021.

Technology-Based Real-Time Visual Feedback in the Education of Singers

Filipa M.B. Lã

Department of Didactics, School Organization and Special Didactics
Faculty of Education, UNED, Madrid, Spain

Learning to sing requires the acquisition of perceptual-motor skills. The development of such skills is notably facilitated when meaningful visual feedback is provided. The current state of voice science, combined with recent technological advances have paved the way for visualization of relevant physiological and acoustical events. Nowadays, non-invasive real-time visual displays of breathing behaviors, subglottal pressure, vibratory patterns and acoustical properties of the voice are available to both teachers and voice students. In this presentation, examples of such displays and the associated technological tools will be demonstrated. For example, the RespTrack system for real-time display of abdominal and ribcage movements will be presented (Johan Stark, Columbi Computers, Sweden). The relationship between breathing behavior, lung volume and subglottal pressure will be discussed, as well as its relevance to the education of singers. Visualization of vocal fold vibratory patterns by electroglottography (EG2-PCX2, Glottal Enterprises, USA) and its application to the training of phonation types or register transitions will be presented. Also, the recently developed FonaDyn freeware will be explored for documenting singers' development (Sten Ternström, Sweden). The usefulness of various spectrographic displays will be discussed. Finally, the possible implementation of all these means in current educational settings will be discussed.

This keynote is an invited summary presentation of the following article:
Lã, F.M.B.; Fiuza, M.B. Real-Time Visual Feedback in Singing Pedagogy: Current Trends and Future Directions. *Appl. Sci.* **2022**, *12*, 10781, doi:10.3390/app122110781. (Open Access)

THE IMPACT OF ROOM ACOUSTICS ON CHORISTERS' PERFORMANCE: FROM REHEARSAL SPACE TO CONCERT HALL

Giorgos Dedousis Areti Andreopoulou Anastasia Georgaki
Laboratory of Music Acoustics & Technology (LabMAT),
National and Kapodistrian University of Athens (NKUA)
gdedousis@music.uoa.gr , a.andreopoulou@music.uoa.gr , georgaki@music.uoa.gr

ABSTRACT

While there has been extensive research on the acoustic quality of various performance spaces and concert halls, studied from the audience perspective, less work has been published on the musicians' on-stage acoustic impression and its impact on musicality and performance quality. On-stage acoustic conditions vary among performance spaces, and, more often than not, between the latter and rehearsal spaces. As a result, there have been studies investigating adaptation mechanisms developed to match specific acoustic conditions during a performance. This paper discusses the potential impact of acoustic mismatches between rehearsal spaces and concert halls on music performance from the perspective of singers and choirs. Based on past research exploring the use of virtual acoustic environments, as a means for investigating this deviation and the way it affects one's performance, a tool is being designed aiming to virtually place users in various spots within a virtual choir on a virtual stage, by augmenting audio recordings with auditory spatialization and room-acoustic cues. Preliminary feedback for the need of this tool along with results from its alpha testing phase are being discussed.

1. INTRODUCTION

There exists extensive research regarding the acoustic experience of the audience in concert halls [1, 2], and the appropriate terminology to describe it [3]. Yet, the on-stage acoustic impression of performers has been studied to a lesser extent. Musicians need to be aware of the sound their instrument produces, the sound coming from the rest of the musicians, and the way the performance space affects this sound. While musicians and audience may share expectations on the qualities of the sound they expect to hear, the auditory perception of the former differs because of the aforementioned interactions between themselves and their instruments, the other musicians, and the performance space [4]. As a result, it has been suggested that the terminology used to describe the musicians' acoustic impressions, while performing, should be different [5].

Studies indicate that on-stage acoustics is an important aspect for conductors, too, as they are responsible for the sound of the orchestra that comes across to the audience. Conductors seem to have their own preferences in acoustics, such as performing in spaces with non-reflective surfaces above the orchestra, while agreeing with both the audience and the musicians in other matters such as the preference for narrow and tall stages [6].

It should be noted that musicians, more often than not, rehearse in spaces that have different acoustic characteristics than performance halls. While rehearsal acoustical needs may differ from those during a performance [7], studies have indicated possible guidelines for these rehearsal areas. The ones that focus on singers, for example, suggest moderately dead rooms, in terms of reverb, which helps the identification of mistakes [8], keeping the reverberation time under one second, as this would be enough to provide performers with assurance and a point of reference while singing [9, 10]. Nevertheless, it has been shown that the acoustic conditions in (music) classrooms where rehearsals take place, are often far from the ideal, for example, vocational classes in Turkey take place in spaces with major deficiencies [11], or renovated spaces in Finland are not suitable for loud music classes [12]. This is to be expected, as very often classrooms are used both for music instruction / lectures and music rehearsals. Hence, "tuning" their acoustics for the needs of both does not seem to be feasible, as speech requires different acoustic treatment than music [13].

Because of all the above, especially when focusing on singing, various strategies have been developed for one's adaptation from the acoustic conditions of a rehearsal space to those of a performance hall [7, 14, 15]. Virtual and Augmented reality technology has been suggested as a means for studying the effects of room acoustic and on-stage acoustic impressions on one's performance, as well as as a tool for helping performers adapt to the acoustic conditions of a performance hall without being physically present in it [16–18]. This study revisits the acoustic mismatch between rehearsal and performance spaces and discusses some of the adaptation mechanisms singers develop to compensate for it. The use of virtual spatial audio technology is explored as a means for helping performers towards the transition from rehearsal to on-stage acoustics, and preliminary feedback on this idea is presented.

2. ROOM ACOUSTICS ADAPTATION

2.1 Instrumentalists' adaptation to room acoustics

Deviations in one's musical performance can often be associated with differences in concert-hall acoustics [19,20]. The ways room acoustics affect musicians during a performance have been studied from various perspectives. It has been reported that musicians try to adapt to the acoustics of a performance space in various ways. For example, when performing in very reverberant rooms, instrumentalists tend to play the staccato notes shorter, so as to render them clearer [21]. In general, larger reverberation times lead solo musicians or smaller ensembles to slower music tempi [22].

Similar changes in music behavior can be induced by virtual acoustic simulations. For example, musicians will adjust staccato-note timing, the sound pressure level, tempo, timbre and amount of vibrato in their music according to the reverberation time of the virtual space they perform in [23,24]. However, whether in real or virtual conditions, while these performance adjustments remain a fact they are not always a conscious choice of the musicians [25].

2.2 Singers' adaptation to room acoustics

Singers and choristers seem to employ similar adaptation strategies to instrumentalists. They tend to feel more satisfied with their performance and enunciate words and music more clearly, in spaces with early reflections with high frequency content [26]. Reverberation time has been found to affect one's comfort level in singing, with early reflections at 40ms marking a threshold above which singing becomes a difficult task [27]. It has been indicated that singers feel more comfortable performing in spaces with moderate reverberation times [28]. "Drier" rooms facilitate rhythmic accuracy and tempo stability.

While Reverberation time has been shown to have minimal impact on one's tonal accuracy [28], it has been found to affect the sound pressure level and dynamic range in one's singing voice [21]. A singer's vocal output level tends to increase as the reverberation time decreases and the performance space becomes more absorbent [14, 21, 29]. The acoustic characteristics of a performance space may also have an impact on a performer's vocal quality and timbre. Singers may choose to increase the resonances of their vocal track [21], or in less reverberant spaces, they may try to move some of their formants higher [29].

In the case of small vocal ensembles the presence of strong early reflections in a room can have a positive impact on music pieces in fast tempi [30]. Reverberation time of a space might even influence the repertoire to be performed, as is implied in another research [21]. Contrary to the findings of [28] the study of [14] provided some evidence that tonal accuracy may differ between a rehearsal space and a performance hall.

As is the case with instrumentalists, the adaptation strategies employed by singers and choristers to compensate for differences in room acoustics, are highly individual [31]. They can be influenced by objective factors such as the auditory and visual feedback singers receive from the performance space, and subjective ones such as their mood or physical condition [32, ch. 16, 17, & 21]. Formal vocal education and performance experience may also have an impact on the strategies employed by singers in any given situation [33].

Despite their highly individual nature [31] and their unconscious selection [25], studies have shown that the adaptation strategies adopted by performers are equivalent in physical and virtual room acoustic conditions [34]. This finding renders virtual rooms a suitable alternative to physical performance spaces for the research of on-stage acoustics and their impact on performance quality and accuracy.

3. PERFORMANCE IN VIRTUAL SPACES

During the COVID-19 pandemic, a lot of music performances took place remotely over the internet, with the aid of virtual reality technology [35], although similar performances had taken place before that, using systems such as the one described in [36]. Such systems have been shown to help performers rehearse and perform without the social and performance anxiety that live concerts sometimes carry [16, 18, 36]. Research on this topic has gained momentum since the pandemic [37], especially because users tend to have positive views of such technologies [16, 38].

It has been suggested that music projects falling under the category of virtual, augmented, mixed, extended reality and augmented virtuality, can be defined as Musical Extended Reality (Musical XR) [37]. Such projects highlight the importance of visual and auditory immersion in remote music interactions [16], while examining the effectiveness of virtual music instruction [38] in different tasks. Virtual Concert Hall (VCH) is a virtual environment combining visual cues and dynamic rendering of binaural sound that was designed to simulate performance spaces [39]. Similarly, VR Rehearse & Perform makes use of a VR headset and headphones to virtually "transfer" musicians to spaces of their choice [18], which was developed based on continuous input from musicians and experts in the field regarding one's needs during a performance. While most such projects rely on binaural sound for auralization, the use of microphones and loudspeakers has also been explored [17].

It is evident that Music XR technology is becoming a popular method for effectively studying the auditory perception and performance practices of musicians in different acoustic conditions, and analyzing the ways musicians interact with each other and with the virtual performance spaces in a musical context [37, 40]. This paper explores the idea of using Music XR technology for assisting choristers become familiar with the on-stage acoustic conditions of a concert hall of their choice, by placing themselves in various positions within a virtual choir, performing the repertoire of their choice. Following previous practices, the exploration of this idea is based on continuous informative feedback from choristers, conductors, and music instructors, used to define common views and user needs.

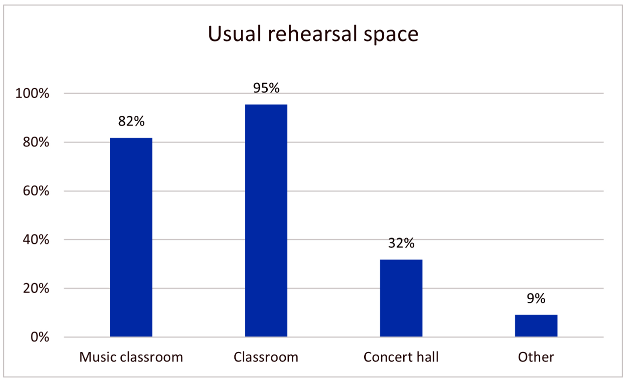


Figure 1. Choristers' usual rehearsal place.

4. CHORISTERS' INQUIRY

Twenty two ($N = 22$) choristers, all undergraduate students of the Department of Music Studies (NKUA), completed a questionnaire regarding the way they perceive the sound they produce as a choir in both rehearsal and performance spaces. Their answers further informed our literature review and identified new concepts and ideas conducive to the development of the research idea discussed. The median age of the participant group was 22 years (min: 18, max: 30 years). The length of time they had spent singing in a choir ranged from a couple of months to 17 years (median: 9 years). 14 of them (64%) had received formal singing lessons (median: 1.75 years).

Figure 1 outlines the variety of room types used for choir rehearsals. As can be seen, rehearsals mainly take place in either regular classrooms, usually used for lectures, or in music classrooms, dedicated to music instruction. Some of the participants have been rehearsing in concert halls and, a small percentage, in other rooms like recording studios (4.5%) and ballet rooms (4.5%). The collected participant responses are in agreement with the literature, according to which, music rehearsals usually take place in multi purpose classrooms [13]. Interestingly though, three out of the four most popular answers to this question concerned rooms that are acoustically treated for music (music classroom, concert hall, recording studio). This rather unexpected finding could possibly be attributed to the conductors' efforts for acoustically sound rehearsals. Further research and a larger sample size could provide more insight into this matter.

The choristers' input on sound perception during a rehearsal and its difference to that perceived during a performance, are illustrated in Figure 2. Participants seem to be aware of the sound they produce during a rehearsal, and are quite content with the acoustic characteristics of the rehearsal space, with only about $1/3$ being moderately satisfied. However, for most of them, the sound they perceive when singing in a performance space, is very to extremely different, to the one they witness during a rehearsal, with less than $1/3$ stating otherwise. These findings are in agreement with the literature [7, 14].

A joined interpretation of the response distributions in Figures 1 and 2, leads to the conclusion that, not only does the sound change between rehearsal and performance spaces, but even in the rare event of rehearsals, taking place in a performance hall, there is still a significant difference between the perceived acoustic impression during a rehearsal and a performance. As this study is a preliminary report, further evidence are necessary to better understand the factors that contribute to this aspect, the ways it may affect the choristers, and the steps that would facilitate the transition between the two.

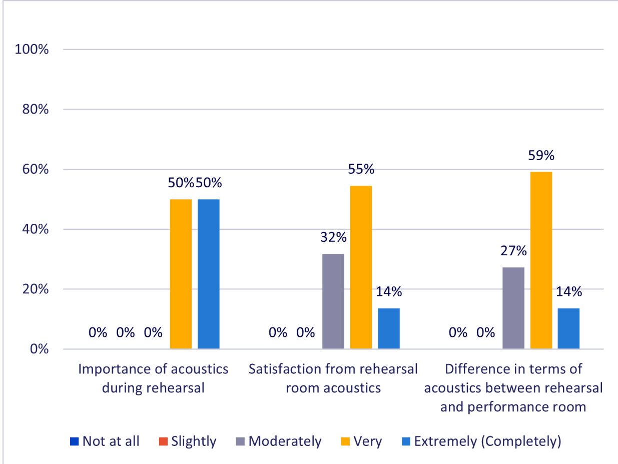


Figure 2. Choristers' responses.

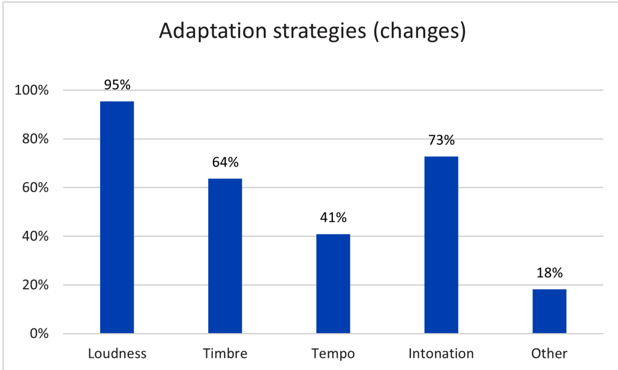


Figure 3. Choristers' adaptation strategies.

The adjustments choristers make to compensate for variations in room acoustic properties is presented in Figure 3. These include changes in loudness, timbre, tempo, intonation, directivity, and other (singing technique, chorister's position in the choir, layout of the choir etc). Changes in loudness, timbre, tempo, and intonation, which were the most frequently suggested answers, also appear in relevant literature [14, 21–24, 28, 29, 31]. Changes in singing technique (suggested by 4.5% of the participants), refer to the methods that have been acquired through singing training (such as breathing), and is where some of the choristers resort to compensate for the difference in room acoustics. This has also been implied in the literature when reported that professional singers tend to choose rooms with different reverberation times than singing students [21]. Changes in the spacing of choristers and height differences among them, often introduced with the use of risers, can also have an impact on one's own perception of sound [41].

Herein, there was also a small percentage of the participants that supported this view by indicating that they would change their own position in the choir ($4,5\%$), or even its complete layout ($4,5\%$).

5. DEVELOPMENT AND ASSESSMENT

The above was a preliminary report on the findings of the way choristers perceive their sound in different rooms, along with adaptation strategies employed to compensate for those. The research idea behind this project is to develop a tool that will support them in the transition between rehearsal and performance spaces, by allowing them to virtually select between different rooms and move between various positions within a virtual choir, performing the repertoire of their choice.

5.1 Development of the first prototype

A prototype tool was designed in the Max/MSP programing environment. In its current early development stage, the auditory content consists of ambisonic recordings of choirs, performing a selected repertoire, captured with three first-order ambisonic microphones which are placed in fixed positions within a choir (far-left front, far-right front, center back). These recordings being conducted in physical concert halls and rehearsal spaces, currently carry the acoustic properties of the rooms they were captured in, with no option for modification. Our future intention is to replace this content with new, recorded in anechoic conditions, and allow users impose on the content the acoustic characteristics of any space they prefer with the use of Room Impulse Responses (RIRs).

For the time-being, the recorded content is binauralized for audition over headphones using a non-individualized, generic Head-Related Transfer Function (HRTF), in a static manner (i.e. no head-tracking is incorporated). The result is a static audio auralization where users can be virtually placed in three fixed positions within the virtual choir, always facing the conductor.

Although, the repertoire of this version of the tool is limited, it covers a wide range of music options from classical music to Greek folk songs and from popular music to children's tunes, allowing for a preliminary assessment of its usability and usefulness by experts in the field.

5.2 Preliminary Evaluation

Upon completion of the development of the first phase of this tool, an open call was made to music teachers and conductors to assess it as part of its alpha-version testing. After being offered a comprehensive introduction on the topic and a thorough presentation of its functionality, participants were asked to use it and complete an evaluation questionnaire. The number of completed questionnaires was eight ($N = 8$). This preliminary study did not touch upon issues of immersion and realism of the experience, due to the uncontrolled listening conditions of the experiment (the audition space and type of headphones used by each participant were not fixed). These will be revisited in the formal evaluation study of the tool. Rather, in this stage we investigated the usability of the tool, the effectiveness of its functionality, as well as its potential use in a real classroom.

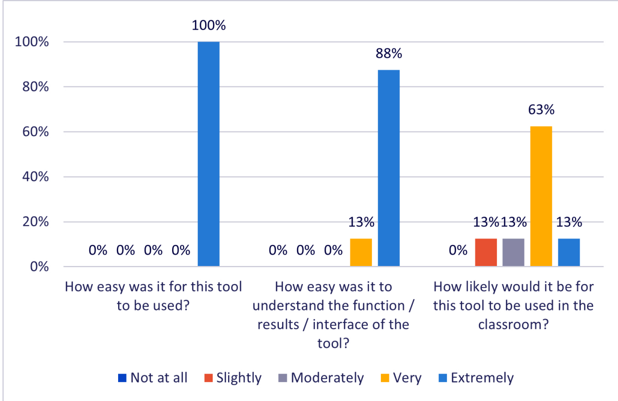


Figure 4. Music teachers' and conductors' responses.

As can be seen in Figure 4, the usability and functionality of the tool have received significantly favorable ratings. As there are only a few choices in the current phase of its development, this was expected, although it provides an indication that its design is on the right path. However, when participants were asked whether this tool could be used in a classroom setting, the responses, although positive, were not as unambiguous. This may be attributed to the tight music curriculum in Greek schools (one hour per week) and conservatories, the latter mainly following a curriculum that was established in 1957 (Greek Law 229/A/1957). The concern that the current curriculum leaves little, if any, time to use such tools was expressed by some of the participants during the introduction.

Participants were also asked to provide any comments / thoughts regarding the tool. The main concepts and ideas that emerged from this open-ended question focused on four main topics. 38% of the participants commented on the possibility of being able to move freely among the virtual environment and hear each chorister / instrument independently, which requires, as stated earlier, a different data collection approach. Two more suggestions concerned the expansion of the functionality and repertoire of the tool, so as to include not only choirs but any kind of ensemble (25%), and to design a user interface that would be as user-friendly and intuitive as possible for the end-user (25%). The interface of the prototype version of the tool can be seen in Figure 5. This version of the tool does not offer a multi-user simultaneous functionality, and this is something that was also expressed as a possible feature by 13% of the participants.Finally, participants commented favorably on the educational perspectives of the tool.

6. CONCLUSIONS AND FUTURE WORK

This paper investigated the potential impact of acoustic mismatches between rehearsal spaces and concert halls on music performance, focusing on singers, and reviewed the adaptation / compensation mechanisms they develop as a result. The role XR Music technology can play on these

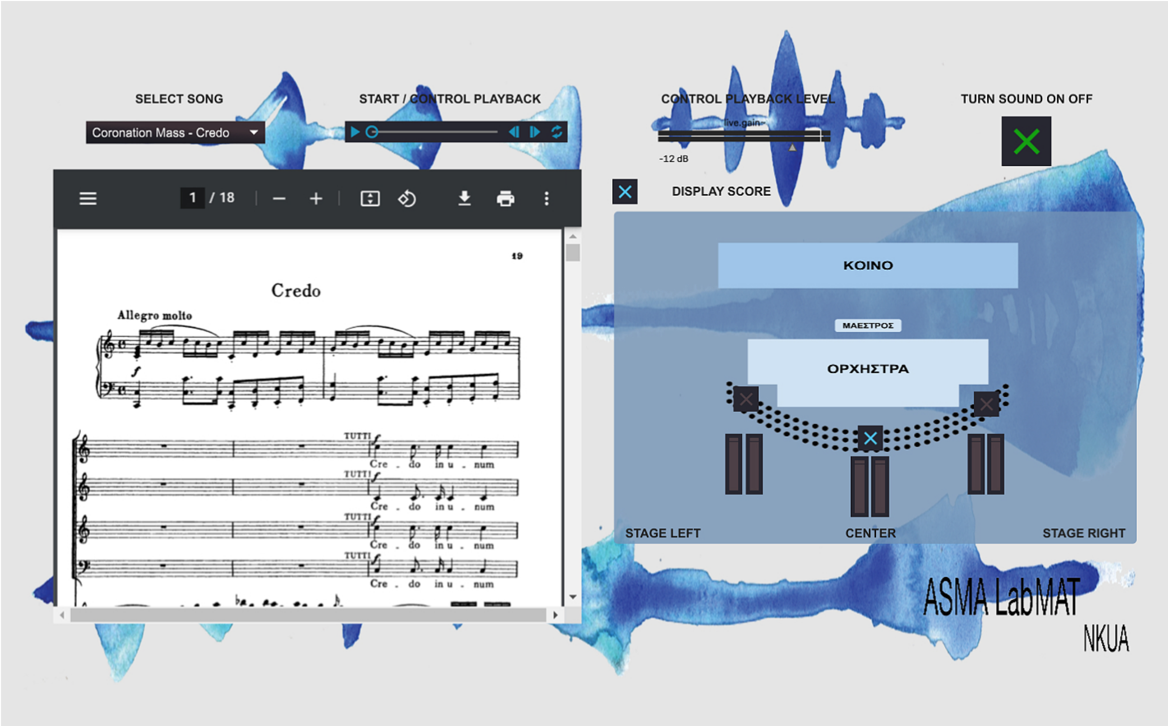


Figure 5. User interface of the preliminary version of the tool.

adjustments was also researched, and was used as a basis for forming and exploring the idea of a tool, which would familiarize choristers with the on-stage acoustic conditions of a concert hall of their choice, by placing them in various spots within a virtual choir, performing a repertoire of their choice.

The aim is to train choristers in acclimatizing to spaces with different acoustic characteristics, while at the same time integrating smoother in a choir, through the use of auditory spatialization. Currently, with the use of headphones and static non - individualized binaural audio content, users can be virtually surrounded by fellow musicians in pre-selected spots within the choir, and experience the impact of position and on-stage acoustics on the perceived auditory feedback.

Future work includes the use of higher-order ambisonics and dynamic binaural rendering, using head-tracking and individualized HRTF data. In doing so, choristers will be able to move freely and in real-time within the choir and, through more simulated concert hall acoustics, place themselves virtually in a wider selection of concert halls. The next step towards this direction concerns data collection (that is music recordings in an anechoic chamber and RIR recordings in various concert halls and rehearsal spaces) so as to simulate room-acoustics as an individual parameter. Also, the repertoire will be increasing both in terms of number of musical pieces and in terms of diversity. As a final step we envision the incorporation of visual cues, which will bring us closer to the development of an XR Music educational tool.

Acknowledgments

Part of this research work was supported by the Hellenic Foundation for Research and Innovation (HFRI) under the "First Call for HFRI Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project Number: HFRI-FM17-3832

7. REFERENCES

[1] L. Beranek, *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*, 2nd ed. New York, NY: Springer, 2004, https://doi.org/10.1007/978-0-387-21636-2. [Online]. Available: http://link.springer.com/10.1007/978-0-387-21636-2

[2] M. Barron, *Auditorium Acoustics and Architectural Design*, 2nd ed. London: Spon Press, Sep. 2009.

[3] A. Kuusinen and T. Lokki, "Wheel of Concert Hall Acoustics," *Acta Acustica united with Acustica*, vol. 103, no. 2, pp. 185–188, Mar. 2017.

[4] J. J. Dammerud, "Stage Acoustics for Symphony Orchestras in Concert Halls," Ph.D. dissertation, University of Bath, Sep. 2009.

[5] Z. S. Kalkandjiev and S. Weinzierl, "An Instrument for Measuring the Perception of Room Acoustics from the Perspective of Musicians: The Stage Acoustic Quality Inventory (STAQI)," in *Fortschritte der Akustik - DAGA 2018 : 44. Jahrestagung für Akustik*, Munich, 2018, pp. 1747 – 1750.

[6] J. Meyer, "Acoustical Demands for the Conductor's Location," *Building Acoustics*, vol. 15, no. 2, pp. 79–94, Jun. 2008, publisher: SAGE Publications

Ltd STM. [Online]. Available: https://doi.org/10.1260/135101008785082506

[7] N. G. Patrick and C. R. Boner, “Acoustics of School-Band Rehearsal Rooms,” *The Journal of the Acoustical Society of America*, vol. 41, no. 1, pp. 215–219, Jan. 1967, publisher: Acoustical Society of America. [Online]. Available: https://asa.scitation.org/doi/abs/10.1121/1.1910321

[8] J. A. Bartle, *Sound Advice: Becoming a Better Children’s Choir Conductor*. Oxford University Press, USA, Mar. 2003.

[9] J. Kramme, “Applications of Acoustical Principles to Selected Problems Arising During Choral Rehearsals,” *The Choral Journal*, vol. 18, no. 7, pp. 5–13, 1978, publisher: American Choral Directors Association. [Online]. Available: https://www.jstor.org/stable/23544944

[10] O. Sinal and S. Yilmazer, “A Comparative Study on Indoor Sound Quality of the Practice Rooms upon Classical Singing Trainees’ Preference,” in *Proceedings of Euronoise 2015*, C. Glorieux, Ed., Maastricht, 2015, pp. 697–702.

[11] M. Gok and C. Sen, “Evaluation of the Physical and Acoustical Competences of Vocational Music Education Institutions in Terms of the Requirements of Music Education,” *International Online Journal of Education and Teaching*, vol. 9, no. 1, pp. 284–307, 2022, publisher: Informascope ERIC Number: EJ1327659. [Online]. Available: https://eric.ed.gov/?id=EJ1327659

[12] H. Koskinen, E. Toppila, and P. Olkinuora, “Facilities for Music Education and Their Acoustical Design,” *International Journal of Occupational Safety and Ergonomics*, vol. 16, no. 1, pp. 93–104, Jan. 2010, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10803548.2010.11076831.

[13] A. Aslan, A. Oktav, and B. Metin, “Determination of the acoustical performance of multi-purpose music classrooms,” *Architectural Science Review*, vol. 0, no. 0, pp. 1–15, Oct. 2022, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00038628.2022.2128711.

[14] K. S. Hom, “The effect of two different rooms on acoustical and perceptual measures of SATB choir sound,” PhD Thesis, University of Kansas, 2013.

[15] L. S.-J. Tsaih, “Soundscape of music reheaarsal in band room,” Ph.D., University of Florida, United States – Florida, 2011, iSBN: 9781303819308. [Online]. Available: https://www.proquest.com/docview/1513246253/abstract/7EDB736BAC59400EPQ/1

[16] L. Aufegger, R. Perkins, D. Wasley, and A. Williamon, “Musicians’ perceptions and experiences of using simulation training to develop performance skills,” *Psychology of Music*, vol. 45, no. 3, pp. 417–431, 2017, publisher: Sage Publications Sage UK: London, England.

[17] E. K. Canfield-Dafilou, E. F. Callery, J. S. Abel, and J. Berger, “A Method for Studying Interactions between Music Performance and Rooms with Real-Time Virtual Acoustics,” in *Audio Engineering Society Convention 146*. Audio Engineering Society, 2019.

[18] S. Ppali, V. Lalioti, B. Branch, C. S. Ang, A. J. Thomas, B. S. Wohl, and A. Covaci, “Keep the VRhythm going: A musician-centred study investigating how Virtual Reality can support creative musical practice,” in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 1–19. [Online]. Available: https://doi.org/10.1145/3491102.3501922

[19] Z. S. Kalkandjiev and S. Weinzierl, “The influence of room acoustics on solo music performance: An empirical case study,” *Acta Acustica united with Acustica*, vol. 99, no. 3, pp. 433–441, 2013.

[20] J. Berg, S. Jullander, P. Sundkvist, and H. Kjekshus, “The Influence of Room Acoustics on Musical Performance and Interpretation - A Pilot Study,” in *Audio Engineering Society Convention 140*. Audio Engineering Society, 2016. [Online]. Available: https://www.aes.org/e-lib/browse.cfm?elib=18282

[21] K. Kittimathaveenan and M. Park, “The influence of stage acoustics on the singers’ performance and perception: A pilot study,” in *Audio Engineering Society Convention 151*. Audio Engineering Society, 2021.

[22] J. Berg, “Concert Hall Acoustics’ Influence on the Tempo of Musical Performances,” in *AES 147th Convention*. New York, USA: Audio Engineering Society, Inc., 2019. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:ltu:diva-77231

[23] K. Kato, K. Ueno, and K. Kawai, “Effect of Room Acoustics on Musicians’ Performance. Part II: Audio Analysis of the Variations in Performed Sound Signals,” *Acta Acustica united with Acustica*, vol. 101, no. 4, pp. 743–759, Jul. 2015.

[24] Z. S. Kalkandjiev and S. Weinzierl, “The influence of room acoustics on solo music performance: An experimental study,” *Psychomusicology*, vol. 25, no. 3, p. 195, 2015, publisher: American Psychological Association.

[25] ——, “Room acoustics viewed from the stage: Solo performers’ adjustments to the acoustical environment,” in *International Symposium on Room Acoustics*, 2013, pp. 9–11.

[26] D. B. Fry, “The singer and the auditorium,” *Journal of Sound and Vibration*, vol. 69, no. 1, pp. 139–142, Mar. 1980. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0022460X8090440X

[27] A. H. Marshall and J. Meyer, "The Directivity and Auditory Impressions of Singers," *Acta Acustica united with Acustica*, vol. 58, no. 3, pp. 130–140, Aug. 1985.

[28] T. Fischinger, K. Frieler, and J. Louhivuori, "Influence of virtual room acoustics on choir singing." *Psychomusicology: Music, Mind, and Brain*, vol. 25, no. 3, p. 208, 2015, publisher: Educational Publishing Foundation.

[29] S. Ternström, "Long-time average spectrum characteristics of different choirs in different rooms," *Voice (UK)*, vol. 2, pp. 55–77, 1993.

[30] D. Noson, S. Sato, H. Sakai, and Y. Ando, "SINGER RESPONSES TO SOUND FIELDS WITH A SIMULATED REFLECTION," *Journal of Sound and Vibration*, vol. 232, no. 1, pp. 39–51, Apr. 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022460X99926846

[31] P. Luizard, J. Steffens, and S. Weinzierl, "Singing in different rooms: Common or individual adaptation patterns to the acoustic conditions?" *The Journal of the Acoustical Society of America*, vol. 147, no. 2, pp. EL132–EL137, 2020, publisher: Acoustical Society of America. [Online]. Available: https://asa.scitation.org/doi/full/10.1121/10.0000715

[32] R. Parncutt and G. McPherson, *The science and psychology of music performance: Creative strategies for teaching and learning*. Oxford University Press, 2002.

[33] O. Sinal and S. Yilmazer, "Effects of perceived singing effort on classical singers' reverberation time preferences towards music practice rooms," *Applied Acoustics*, vol. 136, pp. 132–138, 2018.

[34] P. Luizard, E. Brauer, and S. Weinzierl, "Singing in Physical and Virtual Environments: How Performers Adapt to Room Acoustical Conditions," in *Audio Engineering Society Conference 2019*. Audio Engineering Society, 2019. [Online]. Available: https://www.aes.org/e-lib/browse.cfm?elib=20448

[35] B. Mróz, P. Odya, and B. Kostek, "Creating a Remote Choir Performance Recording Based on an Ambisonic Approach," *Applied Sciences*, vol. 12, no. 7, p. 3316, 2022, number: 7 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2076-3417/12/7/3316

[36] S. Serafin, A. Adjorlu, L. Andersen, and N. Andersen, "Singing in Virtual Reality with the Danish National Children's Choir," in *Perception, Representations, Image, Sound, Music*, ser. Lecture Notes in Computer Science, R. Kronland-Martinet, S. Ystad, and M. Aramaki, Eds. Cham: Springer International Publishing, 2021, pp. 563–574.

[37] L. Turchet, R. Hamilton, and A. Çamci, "Music in Extended Realities," *IEEE Access*, vol. 9, pp. 15 810–15 832, 2021, conference Name: IEEE Access.

[38] K. Pedersen, V. Hulusic, P. Amelidis, and T. Slattery, "Spatialised Audio in a Custom-Built OpenGL-based Ear Training Virtual Environment," *IEEE Computer Graphics and Applications*, vol. 40, no. 5, pp. 67–81, 2020, number: 5. [Online]. Available: https://eprints.bournemouth.ac.uk/34409/

[39] H.-J. Maempel and M. Horn, "The Virtual Concert Hall—A Research Tool for the Experimental Investigation of Audiovisual Room Perception," *International Journal on Stereo & Immersive Media*, vol. 1, no. 1, pp. 78–98, 2017.

[40] J. Brereton, "Music perception and performance in virtual acoustic spaces," in *Body, Sound and Space in Music and Beyond: Multimodal Explorations*, 1st ed., ser. SEMPRE Studies in The Psychology of Music, C. Wöllner, Ed. London: Routledge, 2016, pp. 211–234, num Pages: 24.

[41] J. F. Daugherty, M. L. Grady, and R. C. Coffeen, "Effects of choir spacing and riser step heights on acoustic and perceptual measures of SATB choir sound acquired from four microphone positions in two performance halls," *Journal of Research in Music Education*, vol. 67, no. 3, pp. 355–371, 2019, publisher: SAGE Publications Sage CA: Los Angeles, CA.

Singing Voice Range Profiling Toolbox with real-time interaction and its application to make recording data reusable

Hideki Kawahara
Wakayama University
kawahara@wakayama-u.ac.jp

Ken-Ichi Sakakibara
Health Sciences Univ. Hokkaido
kis@hoku-iryo-u.ac.jp

Kohei Yatabe
Tokyo Univ. Agriculture and Tech.
yatabe@go.tuat.ac.jp

Mitsunori Mizumachi
Kyushu Institute of Technology
mizumach@ecs.kyutech.ac.jp

Tatshya Kitamura
Konan University
t-kitamu@konan-u.ac.jp

ABSTRACT

The Singing Voice Range Profiling Toolbox is a software suite that provides real-time interaction for the profiling of singing voices. It utilizes sound field measurement, microphone calibration, and acoustic characteristics of the recording system to analyze and visualize various vocal parameters. The measurement of the sound recording field and background noise provide information to decide the acceptable distance for optimal performance depending on the directivity pattern and frequency response of microphones. The voice range profiling includes essential parameters such as fundamental frequency ($f_{\rm o}$ [1]), sound pressure level (SPL), cepstral peak prominence (CPP), and EGG-Oq (if EGG is available). In addition, the toolbox provides real-time feedback on the analyzed characteristics, including visualizations of F1 and F2, which are essential and valuable parameters in studying singing voices.

Additionally, the toolbox has facilities for assisting in training and self-learning. The facilities allow users to gain a deeper understanding of the voice range profiling process and improve their skills over time. The Singing Voice Range Profiling Toolbox provides a valuable tool for voice scientists, recording engineers, and singing voice educators, enabling them to make recording data reusable and further advancing the field of voice research.

1. INTRODUCTION

Computing devices today are 10^9 times more powerful than those available a half-century ago [2]. This advancement enables us to provide interactive and real-time tools (including valued voice range profile [3, 4] as a critical component) to assist singing research, training, and self-learning. Although the authors speculate that the rapidly advancing machine learning approaches makes those tools more attractive and valuable, we introduce signal processing-based tools here. Because they are easy to understand and have a considerable body of scientific/practical studies, we focus on this signal-processing-based approach as the first step.

The singing voice results from a specific voice production process and is an acoustic physical entity. Measuring its acoustic attributes is fundamental to starting analysis and providing assisting feedback. Appropriate equipment setting and signal processing procedures are crucial in this stage. Our recent invention of essential algorithms [5, 6] motivated us to develop this toolbox. Those algorithms serve as an infrastructure for a wide range of system analyses (electro acoustic systems, room acoustics, audio coding algorithms, $f_{\rm o}$ estimation algorithms, voice $f_{\rm o}$ response to auditory stimulation, for example) [7–10] and facilitate such tools significantly.

In the following sections, we first introduce motivation and background. Second, we outline effecting factors in voice material acquisition and introduce tools for testing conditions. Third, we introduce analysis and visualization tools of singing voice attributes. Finally, we discuss other available tools and further issues. We placed technical details in Appendix for readability.

2. BACKGROUND

Detailed acoustic analysis of voice attributes requires proper acquisition protocol of voiced sounds [11, 12]. Voice attributes analysis also requires proper terminology established on the scientific foundation [1]. However, preparation for such required conditions is not practically feasible for learners and trainees in everyday life. Note that acoustic attributes for singing are not equally sensitive to the deterioration of recording conditions. We wanted to make voice resources as reusable as possible by reviewing the susceptibility of each attribute on degradations and by providing tools and protocols for the target applications [13]. We extended and applied our essential algorithms [5,6] and found that voice recordings with lossy coding (for example, MPEG-4 AAC [14]) are usable depending on applications and coding details. This finding makes a significant amount of existing materials consisting of voice recording reusable because most use lossy coding. We wanted to provide tools quickly to test materials they were interested in for their applications.

Since 1986, we introduced several interactive tools and algorithms for speech communication research [15–20]. We

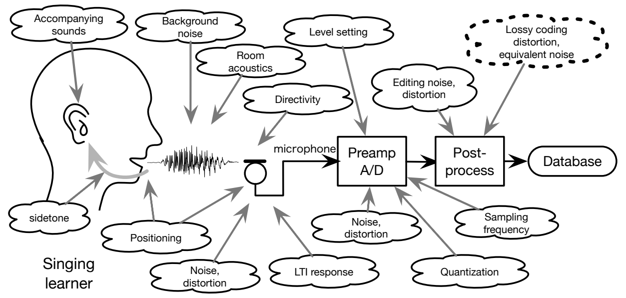


Figure 1. Contributing factors affecting singing voice acquisition.

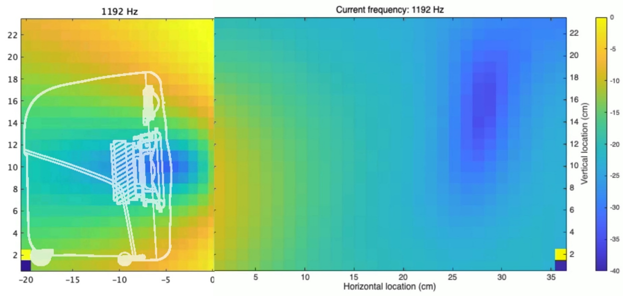


Figure 2. Snapshot of an animation of the sound radiation pattern of each sinusoid (title shows the frequency). The right part shows the sound pressure level, and the left shows the error magnitude by assuming a point sound source (both in dB shown in the colorbar). The wireframe shows the structure of the powered loudspeaker tested.

found they significantly facilitated knowledge acquisition in speech production, perception, analysis, and synthesis. (Recent examples are in [21–24].) Their effectiveness motivated us to integrate these scattered tools and algorithms to develop singing voice research, engineering, education, and learning tools.

Advancements in software environments also facilitated interactive and real-time tools. For example, the scientific computing environment MATLAB [25] provides real-time processing (Audio Toolbox, since 2016) and has a software development environment (Appdesigner, since 2016) for interactive GUI (Graphical User Interface) applications [1]. A combination of computational power and software development advancements motivated us to substantiate tools and protocols we wanted based on our essential signal-processing algorithms and applicatons [5, 6, 8–10].

Finally, the vast range of applicability of our essential algorithms [5,6] and their extensions motivated us to develop this toolbox. The tool designed for voice f_o response measurement to auditory stimulation [8, 22] refines our previous study about three decades ago [26,27]. The tool and its functionality led to the development of an objective evaluation method of f_o estimation algorithms [9]. Application of this evaluation method for tuning our f_o estimation algorithm [28, 29] made it the best algorithm for singing voice study. We want to revise and integrate our previous tools' design and implementation using these essential algorithms [5, 6] and the advancement mentioned above.

3. VOICE ACQUISITION CONDITIONS

Figure 1 illustrates factors affecting voice acquisition conditions. Disturbing noise (background noise, quantization, pre-amplifier noise), linear time-invariant (LTI) deviations (room acoustics, microphone characteristics, directivity and positioning of microphone), signal-dependent deviations (non-linear distortion, distortion by lossy coding), and random and time-varying deviations (temperature, movement, airflow) are sources of acquisition degradations. Note that the acoustic environment (background noise, accompanying sounds, voices, sidetone of own voice, room acoustics) modifies singers' behavior.

[1] 25000 lines of Pascal implemented "Speech Factory" in 1986 [15]. Five hundred lines of MATLAB using "Appdesigner" can do now.

3.1 Tools for preparation

Our algorithms analyze voice acquisition conditions. The first algorithm uses an extended time-stretched-pulse called CAPRICEP (Cascaded All-Pass filters with RandomIzed CEnter frequencies and Phase polarity [5]). It simultaneously measures the LTI, signal-dependent time-invariant, and random and time-varying responses. Impulse response provides room acoustic attributes [30]. It also provides microphone and positioning-related attributes. The second algorithm, "signal safequard [6]" modifies any acoustic contents (for example, speech sounds, music, and alarm) applicable to measure acoustic attributes. The algorithm also provides the same attributes as CAPRICEP-based ones by repetitive measurements. For these attributes to be meaningful, proper calibration is essential.

3.1.1 Calibration

We propose using a deterministic pseudo-band-noise for acoustic calibration instead of a common 1 kHz sinusoid in everyday-life conditions. The proposed calibration signal is a sum of sinusoids with randomized initial phase and frequencies spanning the one kHz octave-band range in a fixed frequency step (for example, one Hz). Figure 2 explains why using a single frequency signal for acoustic calibration is irrelevant in usual recording conditions. Figure 2 is a snapshot of a visualization movie [2] of sound radiation patterns measured using a miniature omnidirectional condenser microphone (DPA-4060) attached to the robot arm (DOBOT Magician) and a sum of sinusoids. Measurement was conducted in a small Japanese room (the room volume is about 20 m^3, and the reverberation time is 0.2 s). Even in the vicinity of the sound source, room acoustic modifies in the order of 10 dB for each sinusoid. Using many (for example, more than a hundred) sinusoids suppresses calibration error to ± 1 dB. We implemented this calibration facility in the proposed tools.

3.1.2 Equipment setting

The LTI response represented by impulse response provides information for assisting equipment setting. Rever-

[2] https://youtu.be/rftqTFH7hh0

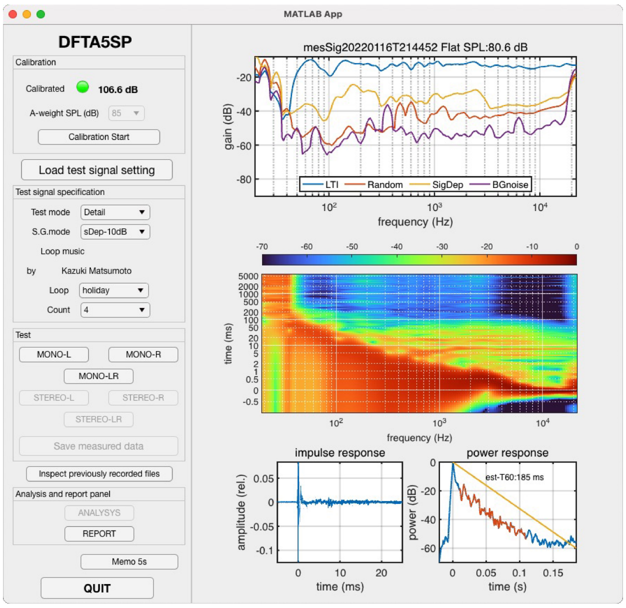


Figure 3. GUI of the sound safeguard-based tool for acoustic attribute measurement.

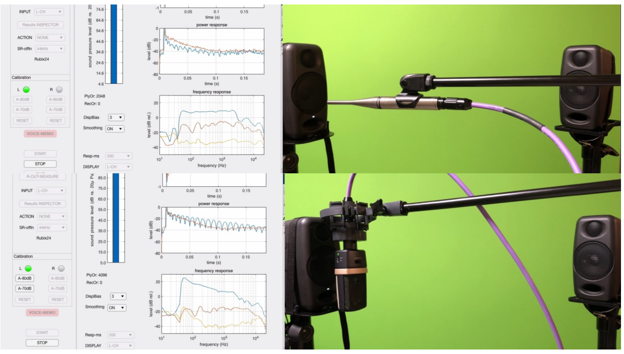


Figure 4. Comparison of snapshots demonstrating the proximity effect of a cardioid pattern directional microphone and non-proximity effect of an omnidirectional microphone.

beration time and critical ratio are such attributes [12]. The frequency response derived from the impulse response is valuable for positioning microphones [11, 12].

Figure 3 shows a snapshot of the sound safeguard-based tool [10]. The left panel has control sub-panels and buttons. The top sub-panel is for calibration. It uses a sum of sinusoids for the calibration signal. The right panel displays acoustic attributes and the impulse response. The top plot shows frequency responses, including the LTI response, and the bottom right shows reverberation time.

Interactive and responsive visualization of these attributes helps appropriately select and position microphones. Figure 4 shows an example. Comparison between the demonstration movie [3] which shows the proximity effect of a directional microphone (AKG C314 by setting cardioid pattern) using older CAPRICEP-based tool [7] and the movie [4] using an omnidirectional microphone (Earthworks M50 which

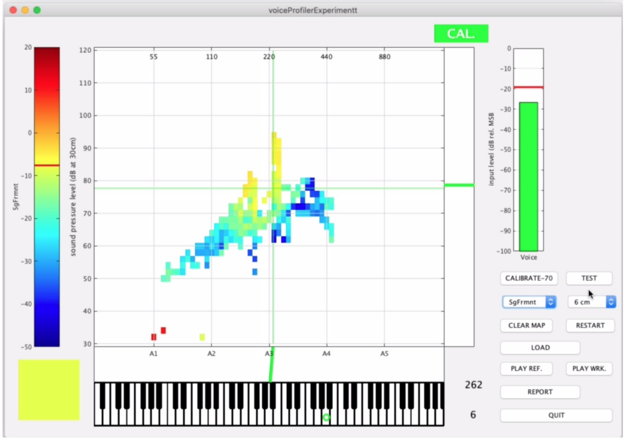


Figure 5. GUI snapshot of a valued voice range profiling tool [24].

has flat response upto 50 kHz) which does not show proximity effect is instructive. Note that we plan to merge the CAPRICEP-based tool into this sound safeguard-based tool because we found that the second essential algorithm's extension is compatible with the CAPRICEP-based algorithm as an efficient implementation of its deconvolution process. There, CAPRICEP provides a family of test signals to the revised tool. Please refer to Appendix for details.

4. TOOLS FOR PROFILING

This section introduces profiling-related tools we developed followed by a prototype integrated tool. References provide details of some of them.

4.1 Real-time valued voice range profiler and f_o-related attributes feedback

Figure 5 shows a snapshot of a valued voice range profiler tool [24]. We implemented it using the "GUIDE" application development environment of MATLAB. Note that "GUIDE" is obsolete since introduction of "Appdesigner." Co-authors (E. Haneishi, a professor and a trainer of music therapy, and K. Hagiwara, a professor and a professional soprano singer) of the reference [24] reported that the piano key at the bottom was not helpful for trainees as visual feedback of trainees' voice f_o. We found some trainees cannot understand instructions such as "raise pitch" or "lower pitch." We needed a more direct and easy-to-understand feedback display. The following tool tried it.

Figure 6 shows a snapshot of a fundamental frequency feedback tool for training pitch awareness [24]. We found that fundamental frequency feedback using a musical note (a red circle) on the staff notation helped trainees understand instructions about voice pitch control. However, the note names next to the left side of the staff notation were less helpful for trainees than the numbers displayed on the right side.

The three left plots represent from top to bottom waveform, f_o trajectory, and pitch salience (Value 1 represents the highest salience). They represent temporal trajectories

[3] https://youtu.be/EPJhCIDC7Zo
[4] https://youtu.be/-9zwVeAz-g0

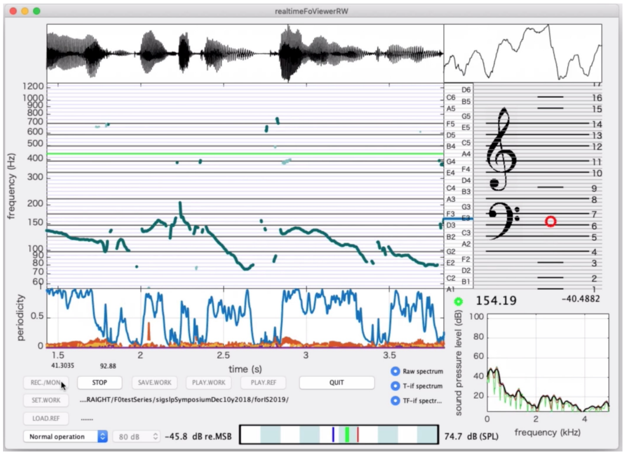


Figure 6. Fundamental frequency feedback tool for training pitch awareness.

and scroll right to the left in real-time. This tool has a playback function to play the recorded trainee's voice and the trainer's pre-recorded voice example corresponding to the displayed portion based on the co-authors' suggestions.

We also designed a real-time visualization tool [24] using wavelet analysis showing amplitude, phase, instantaneous frequency, and group delay with automatic synchronization to the fundamental component. We found this tool too sophisticated even for trainers to use in training.

4.2 Objective evaluation of f_o estimation algorithms

Fundamental frequency f_o is the essential attribute of the singing voice. We have been developing pitch extractors [16, 17, 28, 29, 31, 32] for our previous tools and tested several existing f_o estimation algorithms [33–37]. We needed an objective and more informative evaluation method for f_o estimation algorithms in addition to existing measures. We noticed existing f_o extraction algorithms show non-linear and time-varying (apparently) random behavior, possibly because of the result of fine-tuning to meet the application's requirement. The simultaneous multi-response analysis of the CAPRICEP-based method quantifies that behavior.

We applied the CAPRICEP-based method to generate test signals and analyze LTI, non-linear TI, and random and time-varying attributes [9]. The procedure replaces the human in the test method of voice f_o response to FM auditory stimulation [8, 22] with the test target f_o estimation algorithms.

Figure 7 shows an example of the analysis results. This tool is not real-time and implemented as a live script of MATLAB. We added explanations to a typical output plot for a test signal for 120 Hz carrier f_o with modulation depth 25 cent. This target f_o estimation algorithm uses the instantaneous frequency of the band-pass filter (made from a no-sidelobe Gaussian envelope and a complex exponential impulse response tuned to the target f_o) output.

We reported this method and test results of about 20 f_o estimation algorithms [9]. We tested each f_o estimation algorithm from 80 Hz to 800 Hz f_o in 1/48 octave steps.

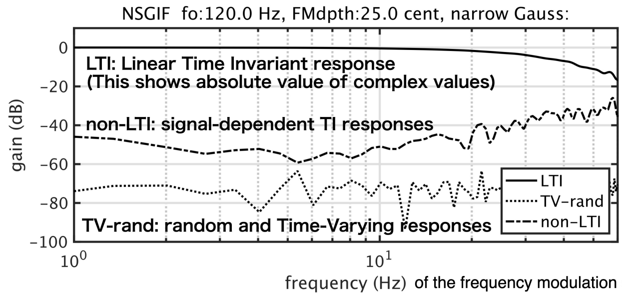


Figure 7. Objective measurement of f_o estimation algorithms.

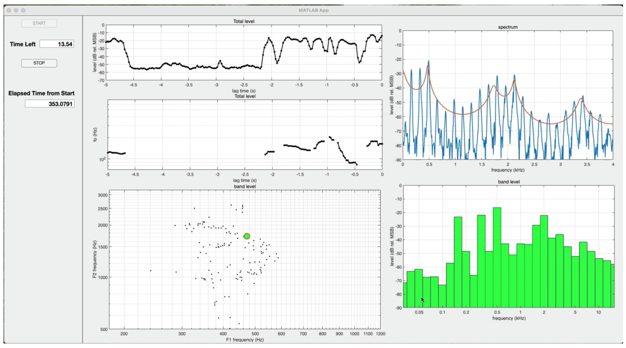


Figure 8. Prototype GUI for F1-F2 feedback.

The accompanying media file [5] of [9] is a visualization movie made from more than 4000 plot results. We found this movie visualization informative to explore and design cost functions for fine-tuning the f_o estimation algorithm to best fit with the desired target applications. We are further revising and simplifying the best-tuned f_o estimation algorithm [9] "NINJALX2" (made from [29]) for the integrated tools we are developing.

Note that this objective evaluation tool does not replace existing evaluation measures of f_o estimation algorithms. Our evaluation tool provides supplemental information which quantifies f_o estimation algorithms in terms of frequency de-modulation accuracy.

4.3 Prototype of revised tool

We are developing the target toolbox by integrating these updated components and representations. We start building prototype tools to test functionality and usability.

Figure 8 shows a snapshot of a skeleton prototype GUI for F1-F2 feedback. The left panel is for control, and the right is for feedback display. The rightmost plots are for real-time feedback. The top plot shows the power spectrum and the LPC spectrum. The bottom plot shows a 1/3 octave band level. The three center plots are for trajectory display. The top shows the RMS level. The middle shows f_o trajectory. The bottom shows an F1-F2 scatter plot with a real-time display of a current voice location in the F1-F2 plane.

We implemented this tool using MATLAB (Appdesigner) by writing less than 400 lines. We will design the next

[5] https://youtu.be/iXnP1tIuVic

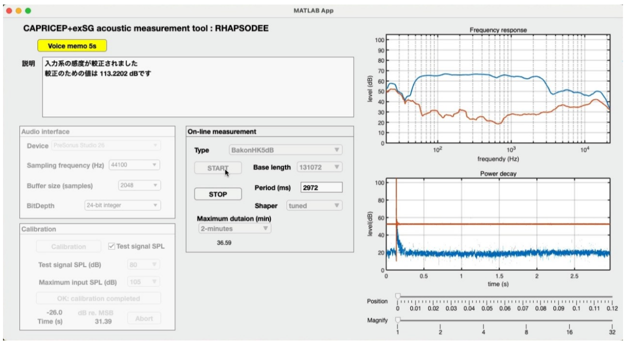


Figure 9. Snapshot of the GUI with a messaging panel. The panel shows instructions in Japanese. This snapshot shows a measurement of a powered loudspeaker using a safeguarded spoken sentence. The blue line in the top right plot shows the LTI response, and the red line shows the disturbance due to background noise. Details are in A.2.

version of the valued voice range profiler by integrating calibration, a staff notation f_o-related attributes feedback, and record and playback functions.

As mentioned in 4.1, our previous tools are not accessible for naïve users. We recently introduced a panel to display instructions for users on what to do next. Authors' and collaborators' students are testing these prototype tools in their projects. Figure 9 shows a snapshot of the GUI with a messaging panel. We will revise prototype tools based on students' feedback and place the initial release of this toolbox by the end of May on our repository [38] and make it open-source.

5. DISCUSSION

Our next goal is to bridge our approach based on the essential algorithms [5,6] and epoch based approach [39,40] for better attribute representation for singing voices. We are confident that the essential algorithms [5,6] are beneficial for many applications. We plan to make these algorithms and their extensions easy to access, understandable and implement for potential users by implementing popular languages and, for example, VST plugins. We know there are many interactive speech tools suitable for education, research, and self-learning (for example, refer [41,42]). We hope our tool also serves as a beneficial tool.

6. CONCLUSIONS

This paper introduced "Singing Voice Range Profiling Toolbox". It is a software suite that provides real-time interaction for the valued voice range profiling of singing voices. It utilizes sound field measurement, microphone calibration, and acoustic characteristics of the recording system to analyze and visualize various vocal parameters. The MATLAB implementation of the toolbox is available in our open-source repository [38].

Acknowledgments

First, we appreciate anonymous reviewers for constructing comments, which helped revise this article more accessible to potential readers. This research is supported by Kakenhi (the Grant in Aids for Scientific Research) 20H00291, 21H00497, 21H01596 by JSPS (The Japan Society for the Promotion of Science). We appreciate E. Haneishi and K. Hagiwara's valuable comments and feedback on profiling and pitch awareness training tools. We appreciate T. Matsui and L. Jiahui for their valuable feedback and discussions on voice f_o experiment and tools. Finally, we appreciate M. Morise's discussions and inspirations on voice analysis and source representations.

7. REFERENCES

[1] I. Titze, R. Baken, K. Bozeman, S. Granqvist, N. Henrich, C. Herbst, D. M. Howard, D. Kaelin, R. Kent, M. Kob *et al.*, "Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization," *J. Acoustical Society of America*, vol. 137, no. 5, pp. 3005–3007, 2015.

[2] C. E. Leiserson, N. C. Thompson, J. S. Emer, B. C. Kuszmaul, B. W. Lampson, D. Sanchez, and T. B. Schardl, "There's plenty of room at the top: What will drive computer performance after moore's law?" *Science*, vol. 368, no. 6495, p. eaam9744, 2020.

[3] P. Pabon, "Mapping individual voice quality over the voice range: the meeasurement paradigm of the voice range profile," Ph.D. dissertation, KTH, 2018.

[4] P. Pabon and S. Ternström, "Feature maps of the acoustic spectrum of the voice," *J. Voice*, vol. 34, no. 1, pp. 161.E1–161.E26, 2020.

[5] H. Kawahara and K. Yatabe, "Cascaded all-pass filters with randomized center frequencies and phase polarity for acoustic and speech measurement and data augmentation," *Proc. ICASSP*, pp. 306–310, 2021.

[6] ——, "Safeguarding test signals for acoustic measurement using arbitrary sounds: Measuring impulse response by playing music," *Acoustical Science and Technology*, vol. 43, no. 3, pp. 209–212, 2022.

[7] H. Kawahara, K. Yatabe, K.-I. Sakakibara, M. Mizumachi, M. Morise, H. Banno, and T. Irino, "Interactive and Real-Time Acoustic Measurement Tools for Speech Data Acquisition and Presentation: Application of an Extended Member of Time Stretched Pulses," in *Proc. Interspeech Show and Tell*, 2021, pp. 4853–4854.

[8] H. Kawahara, T. Matsui, K. Yatabe, K. I. Sakakibara, M. Tsuzaki, M. Morise, and T. Irino, "Implementation of interactive tools for investigating fundamental frequency response of voiced sounds to auditory stimulation," in *Proc. APSIPA ASC*, 2021, pp. 897–903.

[9] H. Kawahara, K. Yatabe, K.-I. Sakakibara, T. Kitamura, H. Banno, and M. Morise, "An objective test tool for pitch extractors' response attributes," in *Proc. Interspeech*, 2022, pp. 659–663, (Comparison movie of 16 pitch extractors: https://youtu.be/iXnP1tIuVic).

[10] H. Kawahara, K.-I. Sakakibara, M. Mizumachi, M. Morise, and H. Banno, "Comparative measurement of headphones using new test signals and side tones while voicing," *Proc. Auditory Res. Meeting*, vol. 52, no. 8 H-2022-104, pp. 575–580, 2022.

[11] J. G. Švec and S. Granqvist, "Guidelines for selecting microphones for human voice production research," *American J. Speech-Language Pathology*, vol. 19, no. 4, pp. 356–368, 2010.

[12] R. R. Patel, S. N. Awan, J. Barkmeier-Kraemer, M. Courey, D. Deliyski, T. Eadie, D. Paul, J. G. Švec, and R. Hillman, "Recommended protocols for instrumental assessment of voice: American speech-language-hearing association expert panel to develop a protocol for instrumental assessment of vocal function," *American J. Speech-Language Pathology*, vol. 27, no. 3, pp. 887–905, 2018.

[13] K.-I. Sakakibara, H. Kawahara, and M. Mizumachi, "Recommended protocols for acoustic recording of speech data for high reusability," *J. Acoustical Society of Japan*, vol. 76, no. 6, pp. 343–350, 2020, [in Japanese].

[14] ISO/IEC 14496-3:2019, *Information technology – Coding of audio-visual objects – Part 3: Audio*, ISO, 2019–12.

[15] NTT Advance Technology Inc., "Speech factory," 1989, (Speech acquisition, editing, and analysis tool with interactive GUI implemented with 25000 lines of Pascal code. Originally designed and coded by Hideki Kawahara in 1986. The original name was "Spark.").

[16] H. Kawahara, I. Masuda-Katsuse, and A. de Cheveigne, "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction," *Speech Communication*, vol. 27, no. 3-4, pp. 187–207, 1999.

[17] H. Kawahara, M. Morise, T. Takahashi, R. Nisimura, T. Irino, and H. Banno, "TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0 and aperiodicity estimation," in *ICASSP 2008*, Las Vegas, 2008, pp. 3933–3936.

[18] H. Kawahara, "SparkNG: Interactive MATLAB tools for introduction to speech production, perception and processing fundamentals and application of the aliasing-free LF model component," *Proc. Interspeech Show and Tell*, pp. 1180–1181, 2016.

[19] M. Morise, F. Yokomori, and K. Ozawa, "WORLD: A vocoder-based high-quality speech synthesis system for real-time applications," *IEICE Transactions on Information and Systems*, vol. 99, pp. 1877–1884, 2016.

[20] V. G. Skuk and H. Kawahara, "Voice Morphing," in *The Oxford Handbook of Voice Perception*, S. Frühholz and P. Belin, Eds. Oxford University Press, 12 2018, pp. 685–706.

[21] T. Matsui, T. Irino, R. Uemura, K. Yamamoto, H. Kawahara, and R. D. Patterson, "Modelling speaker-size discrimination with voiced and unvoiced speech sounds based on the effect of spectral lift," *Speech Communication*, vol. 136, pp. 23–41, 2022.

[22] H. Kawahara, T. Matsui, K. Yatabe, K.-I. Sakakibara, M. Tsuzaki, M. Morise, and T. Irino, "Mixture of Orthogonal Sequences Made from Extended Time-Stretched Pulses Enables Measurement of Involuntary Voice Fundamental Frequency Response to Pitch Perturbation," in *Proc. Interspeech*, 2021, pp. 3206–3210.

[23] H. Terasawa, K. Wakasa, H. Kawahara, and K.-I. Sakakibara, "Investigating the Physiological and Acoustic Contrasts Between Choral and Operatic Singing," in *Proc. Interspeech*, 2019, pp. 2025–2029.

[24] H. Kawahara, K.-I. Sakakibara, E. Haneishi, and K. Hagiwara, "Real-time and interactive tools for vocal training based on an analytic signal with a cosine series envelope," in *Proc. APSIPA ASC*, 2019, pp. 907–910.

[25] *MATLAB version 9.13.0.2126072 (R2022b) Update 3*, The Mathworks, Inc., Natick, Massachusetts, 2022.

[26] H. Kawahara, "Interactions between speech production and perception under auditory feedback perturbations on fundamental frequencies," *J. Acoust. Soc. Jpn. (E)*, vol. 15, no. 3, pp. 201–202, 1994.

[27] H. Kawahara, H. Kato, and J. C. Williams, "Effects of auditory feedback on F0 trajectory generation," in *Proc. Int. Conf. Spok. Lang. Process. (ICSLP)*, vol. 1, 1996, pp. 287–290.

[28] H. Kawahara, Y. Agiomyrgiannakis, and H. Zen, "Using instantaneous frequency and aperiodicity detection to estimate F0 for high-quality speech synthesis," in *9th ISCA Speech Synthesis Workshop*, 2016, pp. 221–228.

[29] H. Kawahara, "Application of time-frequency representations of aperiodicity and instantaneous frequency for detailed analysis of filled pauses," *J. Phonetic Society of Japan*, vol. 21, no. 3, pp. 63–73, 2017.

[30] ISO 3382-2:2008, *Acoustics – Measurement of room acoustic parameters – Part 2: Reverberation time in ordinary rooms*, ISO, 2008–06.

[31] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1917–1930, 2002.

[32] H. Kawahara, A. de Cheveigné, H. Banno, T. Takahashi, and T. Irino, "Nearly defect-free F0 trajectory extraction for expressive speech modifications based on STRAIGHT," in *Proc. Interspeech*, 2005, pp. 537–540.

[33] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in *Proc. ICPhS*, vol. 17, no. 1193, 1993, pp. 97–110.

[34] F. Eyben, M. Wöllmer, and B. Schuller, "openSMILE – the Munich versatile and fast open-source audio feature extractor," in *Proc. 18th ACM Multimedia*, 2010, pp. 1459–1462.

[35] D. Talkin, "REAPER: robust epoch and pitch estimator," 2015, GitHub: https://github.com/google/REAPER.

[36] M. Morise, "Harvest: A high-performance fundamental frequency estimator from speech signals," in *Proc. Interspeech*, 2017, pp. 2321–2325.

[37] J. W. Kim, J. Salamon, P. Li, and J. P. Bello, "CREPE: A convolutional representation for pitch estimation," in *Proc. ICASSP*, 2018, pp. 161–165.

[38] H. Kawahara, "GitHub repository for speech and hearing research/education tools," 2023, (retrieved 31 Jan. 2023). [Online]. Available: https://github.com/HidekiKawahara

[39] B. Yegnanarayana, "Analysis of phase derivatives of speech signals," *J. Acoustical Society of America*, vol. 152, no. 3, pp. 1721–1736, 2022.

[40] S. R. Kadiri, P. Alku, and B. Yegnanarayana, "Analysis of instantaneous frequency components of speech signals for epoch extraction," *Computer Speech and Language*, vol. 78, p. 101443, 2023.

[41] M. Huckvale, "Web tools for speech," INTERNET INSTITUTE FOR SPEECH AND HEARING, 2022, https://www.speechandhearing.net/laboratory/webtools.php (visit:26/Dec./2022).

[42] N. Thapen, "Pink trombone v1.1," Web applicatoin, 2017, https://dood.al/pinktrombone/ (visit:26/Dec./2022).

[43] R. H. Lyon, "Statistics of combined sine waves," *J. Acoust. Soc. Am.*, vol. 48, no. 1B, pp. 145–149, 1970.

A. ESSENTIAL ALGORITHMS

A.1 Algorithm 1: CAPRICEP

This section is a revised version of an excerpt from [5].

The transfer function $H_C(z)$ of the cascaded all-pass filters is the product of constituent transfer function $H_k(z)$ which is an all-pass filter with a center frequency f_k.

$$H_C(z) = \prod_{k=1}^{K} H_k(z), \tag{1}$$

where K represents the total number of cascaded filters. Because the absolute value of the transfer function of an all-pass filter is a frequency-independent constant value, the absolute value of $H_C(z)$ is a frequency-independent constant. The inverse z-transform of $H_C(z)$ provides an impulse response.

A systematic procedure determines each center frequency f_k using two set of random numbers $r_1[k]$ and $r_2[k]$. We introduce random numbers because each filter response corresponds to a damped sinusoid [6]. Intuitively, the sum of many sinusoids with the same amplitude with the random phase yields items that behave like samples from normal distribution [43]. We hope frequency randomization yields noise-like behavior of the impulse response of $H_C(z)$.

The following equation is an implementation of the systematic procedure of filter assignment.

$$f_k = r_1[k]\, F_d \sum_{k=1}^{K} r_2[k], \tag{2}$$

where F_d represents the average interval between aligned center frequencies (*i.e.* $f_{k+1} > f_k$). The random number $r_1[k]$ has the value 1 and -1 with equal probability. Note that $r_1[k] = -1$ makes $H_k(z)$ anti-causal and makes its impulse response time-reversed. This randomization is for design flexibility described in the next section. Without introducing $r_1[k]$, the response is causal and does not provide flexible design.

The random number $r_2[k]$ obeys a probability density distribution function (pdf) $g(x)$ defined on $[0, 1]$. We do not randomize f_k directly.

We introduce the Beta distribution for $r_2[k]$ because it is a simple adjustable distribution having two design parameters. The following is the definition of the Beta distribution.

$$b(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}, \tag{3}$$

where $\Gamma(\alpha)$ is Gamma function. When $\alpha \equiv \beta$, $b(x)$ seamlessly deforms from concave to convex via uniform distribution with α.

For the parameter b_k that represents the group delay width of the all-pass filter, we make it proportional to F_d using a coefficient c_{mag}. In short, we set $b_k = c_{\mathrm{mag}} F_d$ (at least for the first time). In summary, $H_C(z)$ has three design parameters, F_d, c_{mag}, and α. This additional parameter α enables us to design the shape of the waveform distribution. Let $h_c[n]$ represent the impulse response of $H_C(z)$. Inverse discrete Fourier Transform of $H_C(z)$ yields $h_c[n]$, a unit-CAPRICEP.

Figure 10 shows a schematic diagram of simultaneous multiple attributes measurement. Adding three different unit-CAPRICEPs using three orthogonal sequences provides a unit mixture signal. Repeatedly overlap-and-adding the unit mixture provides a periodic test signal. The output of the target system to the test signal input is also a periodic signal after the second cycle of the period. We apply the pulse recovery procedure to the output and demixing yields three impulse responses. Differences between these responses provide signal-dependent and time-invariant attributes. Impulse response differences calcu-

[6] This response is for IIR-type all-pass filter implementation. Another type of all-pass filter has a different response shape.

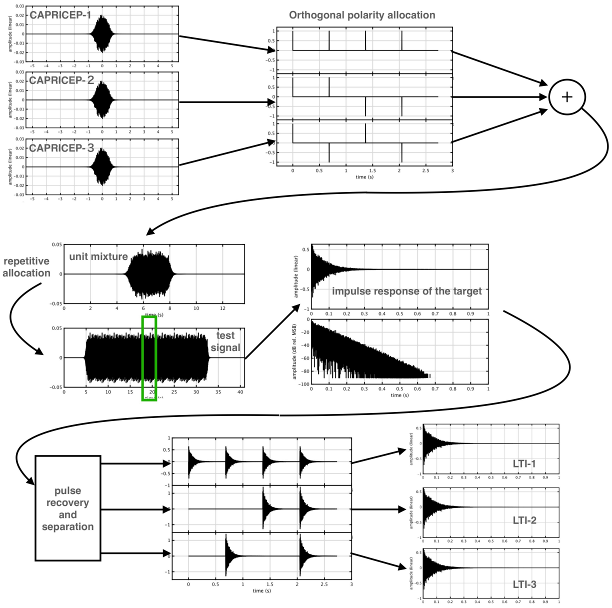


Figure 10. Simultaneous multiple attributes measurement scheme.

lated from different (repetition) periods provide random and time-varying attributes.

A.2 Algorithm 2: Safeguarded signal

This is a revised excerpt form [6]. Let $x[n]$ be a periodic discrete-time signal with a period L. Convolution of $x[n]$ and the impulse response $h[n]$ of the target system yields the output $y[n]$. Because the signal is periodic, the DFT (Discrete Fourier transform) of $x[n]$ and $y[n]$ segments (their length is L) are invariant other than the phase rotation proportional to frequency. Let $X[k]$ and $Y[k]$ represent their DFT, where k, $(k = 0, \ldots, L-1)$, is the discrete frequency. Then, the ratio $Y[k]/X[k]$ is independent of the location of the segment. This ratio agrees with the DFT $H[k]$ of the impulse response $h[n]$, where $X[k] \neq 0$ for all k values is the condition of this relation to provide physically meaningful results.

However, this simple solution is sensitive to noise when the absolute value $|X[k]|$ is very small relative to absolute values $|H[k]|$ of other k values. We propose to limit the absolute value $|X[k]|$ to have larger value than a threshold [7]. We use the following equation to derive the DFT $X_{\mathrm{s}}[m]$ of the safeguarded signal $\tilde{x}_{\mathrm{s}}[n]$.

$$X_{\mathrm{s}}[k] = \begin{cases} \dfrac{\theta_L[k]X[k]}{|X[k]|} & \text{for} \quad 0 < |X[k]| < \theta_L[k] \\ X[k] & \qquad\;\; \theta_L \leq |X[k]| \end{cases}, \quad (4)$$

where we set $X_{\mathrm{s}}[k] = \theta_L[k]$ when $X[k] = 0$. Then, we derive the safeguarded transfer function $H_{\mathrm{s}}[k]$ as follows.

$$H_{\mathrm{s}}[k] = \frac{Y_{\mathrm{s}}[k]}{X_{\mathrm{s}}[k]}, \quad (5)$$

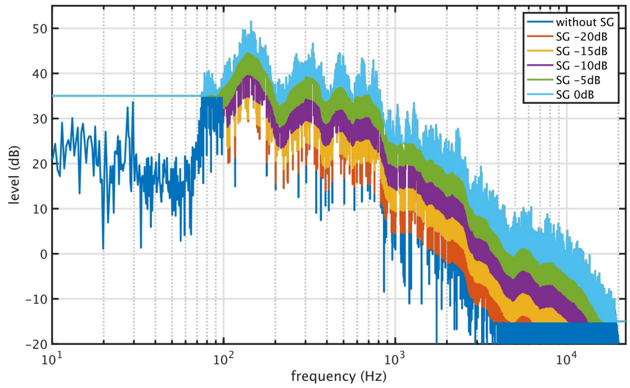


Figure 11. Absolute values of the original (a Japanese sentence /bakuon ga ginsekai no kougeN ni hirogaru/ spoken by a male speaker) and safeguarded speech samples. The signal period is 2^{17} samples at 44100 Hz. Frequency-dependent thresholding uses a smoothed power spectrum with 1/3 octave width as a reference.

where $Y_{\mathrm{s}}[k]$ represents the DFT of the output of the target system for periodic test signal $\tilde{x}_{\mathrm{s}}[n]$. Because the safeguarded signal $\tilde{x}_{\mathrm{s}}[n]$ is periodic, we can make *the safeguarded test signal* for acoustic measurement by concatenating it as many times as required.

Figure 11 shows example of safeguarding. The level represents the absolute values of discrete Fourier transform of a whole sentence length samples (the original and safeguarded ones).

A.3 DFT-based revision of multiple attributes

The original implementation of the CAPRICEP-based multi-attribute measurement used an intricate procedure for canceling crosscorrelation between unit-CAPRICEPs. A DFT-based procedure replaces the original procedure by using the binary sequence for designing the test signal as the target signal. It is conceptually an extension of the safeguarding procedure.

B. RADIATION PATTERN MESSURMENT

We used an education-purpose robot arm (DOBOT Magician) for measuring the sound radiation pattern from a loudspeaker. A movie [8] shows an example measurement of radiation from an iPhone. This measurement also uses a DFT-based procedure.

[7] In actual implementation we use frequency dependent threshold $\theta_L[k]$ using power spectra of the original signal and the background noise. We also set the minimum level and low frequency limit for $\theta_L[k]$.

[8] https://youtu.be/avY7HAFX-50

Vocal-tract impedance at the mouth – from 1995 to today
A tribute to Joe Wolfe and John Smith

Malte Kob
Erich Thienhaus Institute
HfM Detmold, Detmold, Germany
& InstAS, mdw Vienna, Austria
malte.kob@hfm-detmold.de

Jithin Thilakan
Erich Thienhaus Institute
HfM Detmold, Detmold, Germany
jithin.thilakan@hfm-detmold.de

Nathalie Henrich Bernardoni
Univ. Grenoble Alpes
CNRS, Grenoble INP
GIPSA-lab, Grenoble, France
nathalie.henrich@gipsa-lab.fr

ABSTRACT

The acoustic output signal of orators and singers contains information from the voice source as well as from the articulatory organs. The extraction of formants to achieve information about voice timbre and vocal-tract configuration is among the essential methods in speech and language processing as well as in singing-voice research. However, the validity of formant analysis is limited by voice features and analysis parameters. The measurement of the acoustic impedance of resonating structures allows one to obtain all relevant information from their resonances. The design of a method for impedance measurement at the human lips by Joe Wolfe and John Smith dates back to 1995, and since then variations of the concept have been applied in fundamental singing-voice research (high soprano and tenor voices, overtone singing), as well as for investigations in other domains such as voice rehabilitation. This review article collects methods and applications over the last 25 years, giving reference to the many colleagues and students who have contributed with theses and supporting work. The presentation is accompanied by a demonstration of the method.

1. INTRODUCTION

An innovative method for measuring wind-instrument and vocal-tract impedances was developed by Prof. Joe Wolfe and Prof. John Smith starting from 1995 [1, 2]. In the case of speech and singing, it allows the measurement of vocal-tract acoustical resonance properties without making any a priori assumptions about the glottal source. Earlier work of Osamu Fujimura and Jan Lindqvist has investigated the vocal tract transfer function [3] using external swept sine excitation at the larynx. This early research already revealed the significance of glottis, epiglottis and velum configuration on the resonance characteristics of the vocal-tract. In the last decade the investigation of the vocal tract acoustics using various other methods has been facilitated by the availability of MRI-based data and physical models produced with 3D printers [4, 5]. Their results demonstrate a high comparability of resonance measurements using external excitation and formant frequencies derived from human voice signals.

The method presented here excites the vocal-tract externally with a sound wave. The development and applications of this method are detailed below.

The excitation of the vocal-tract using a broadband signal such as noise or a swept sine provides a high frequency resolution of the measurements in contrast to formant analysis methods that need to use the line spectrum of the voiced signal which – in case of high-pitched vowels – does not provide sufficient energy between the partials for an accurate formant analysis. On the other hand the determination of formants from the voice signal provides an average estimation of the vocal-tract's filter function over time which corresponds to our perception of the filtered voice signal.

When the impedance at the mouth is measured during phonation the resulting signal not only contains the impedance function but also the voice signal. For a clear presentation of the impedance function the voice signal must be separated or suppressed to obtain resonance values that correspond to the formants during phonation.

Alternatively, the impedance can also be measured without phonation at all, resulting in undisturbed resonances. However, the glottal status of such measurements can be anything between open (higher resonances than average during phonation) and closed (lower resonances), resulting in a quite large variability of the impedance function and the derived resonance values [6]. Instructions to subjects to keep the glottis open or closed during measurements without phonation can help to define the glottal status but not necessarily provide the desired conditions. Trained subjects, however, can voluntarily open and close their glottis which provides valuable insight to the range of the vocal tract resonances for different open quotients. The position of the epiglottis can vary as well and is difficult to control voluntarily. However, the epiglottis movement would not have an equally fundamental effect on the resonance structure as the glottal status unless it is tilted back completely (as during swallowing). Further limitations of an external excitation at the lips of a speaker/singer can be the surprising sound of a noise or swept sine signal that might cause articulatory changes just before or during to the impedance measurement. To summarise, the impedance function derived at the mouth end of the vocal-tract can not provide the same information as the analysis of a voice signal but

Figure 1. VTMI device

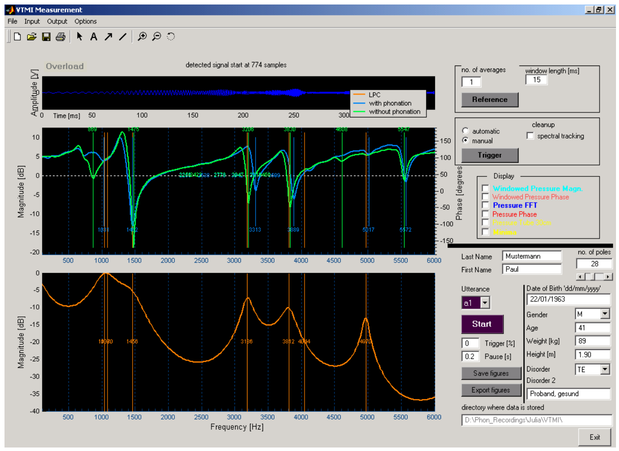


Figure 2. Applet for the VTMI device, impedance curves (top) and LPC curves (bottom) shown simultaneously for vowel /a:/

allows insight into articulatory phenomena that are hard to obtain otherwise.

The method to measure the vocal-tract impedance was first developed for the study of musical instruments [1, 7] but also soon and successfully applied to the study of vowel formants in speech [8, 9] including applications to training of phonation as well as to the investigation of singing [10, 11]. We will first present the devices and their evolution in Section 2. Their application to the analysis of singing voice in different styles will then be reviewed. The nomenclature in this text follows the proposal given in [12].

2. METHOD

The measurement of the acoustic impedance requires the determination of two basic properties of the sound field: acoustic pressure $p(t)$ & acoustic velocity $v(t)$. The acoustic impedance $Z(f)$ is then determined by the quotient of the RMS values of both:

$$Z(f) = \frac{\mathrm{FFT}(\tilde{p}(t))}{\mathrm{FFT}(\tilde{v}(t))}. \tag{1}$$

The acquisition of acoustic pressure is far easier since sensors for acoustic velocity are, due to their small size, expensive, sensitive to damage, and of limited accuracy. Despite these problems, a prototype that captures both sound pressure and velocity has been developed in 2002. The installation of the sensors in a prototype is shown in Fig. 1. A graphic user interface for parallel evaluation of curves obtained by linear predictive coding (LPC) and vocal tract impedance at the mouth (VTMI) was developed in Matlab ([13], Fig. 2). Within the dissertation project of Julia Stoffers [14] the evaluation of the impedance curves was performed with respect to several parameters which were derived from the normalised magnitude of the impedance.

Candidates for the assessment of values for amplitude and bandwidth of each resonance were the steepness and the difference between maximum and minimum. It was found that an absolute amplitude or bandwidth was difficult to obtain since these values do not seem to be constant with frequency [6].

2.1 One or two sensors?

Another device was developed at RWTH Aachen University in 2003 to provide both acoustic pressure and velocity values using miniature sensors in a small, portable set-up that allows mobile measurements, see Fig 3. The KE4-211-2 capsules by Sennheiser were used for pressure, and prototypes by Microflown were used for velocity. In Fig. 4 the set-up (right) and the velocity sensor (left) are shown.

By acoustic treatment of the propagation of the acoustic wave from the exciter to the exit of the measurement device, the acoustic velocity can be held approximately constant. This condition can be achieved by guiding the sound wave through a capillary wave guide or an inverse horn [15]. The use of a horn and additional damping material greatly reduces the feedback of an external impedance on the loudspeaker membrane. A constant sound velocity supply can be assumed. As a consequence, the relative impedance can be determined by simply measuring sound pressure $p(t)$ instead of also measuring sound velocity $v(t)$.

2.2 Calibration

The combined acquisition of sound pressure $p(t)$ and velocity $v(t)$ allows the direct calculation of the acoustic impedance by Eq. 1. If a change of the acoustic impedance is sufficient for characterisation of the resonances, and no absolute values are required, the measurement device can be calibrated against a constant impedance — such as the impedance Z_0 of the free field around the sensor, or the impedance of the closed mouth. Resonances of the vocal tract with an open mouth would then be shown as a deviation of the measured impedance from Z_0.

Practically, the reference Z_0 is determined just prior to the measurement at the mouth and stored. After the acquisition of the sound pressure from the calibration ($p_0(t)$)and

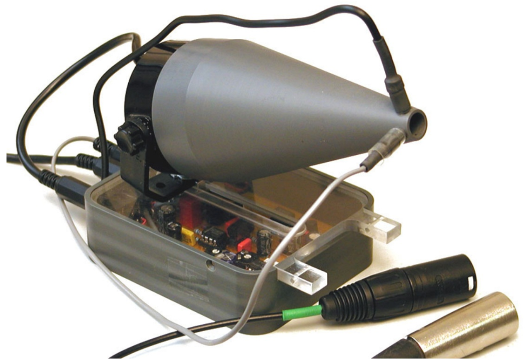


Figure 3. Minature VTMI set-up

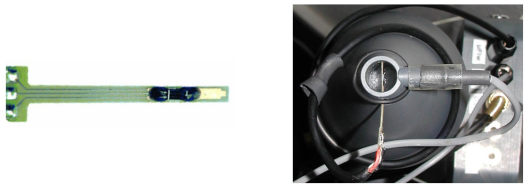


Figure 4. Sensors of miniature VTMI set-up

the excitation at the mouth ($p_m(t)$), and subsequent Fourier transform, the relative impedance $Z_{\text{rel}}(f)$ of an ideal current source having uniform velocity is then calculated by the quotient of both measurements:

$$Z_{\text{rel}}(f) = \frac{Z_m}{Z_0} = \frac{\text{FFT}(\tilde{p}_m(t))}{\text{FFT}(\tilde{p}_0(t))}. \tag{2}$$

$Z_{\text{rel}}(f)$ is either visualised by its absolute value or phase (see Figs. 2 and 7-9).

2.3 Signal extraction

Methods for extraction of the acoustic impedance from a measurement during phonation have been developed from the beginning. Whereas early concepts used manual extraction of the resonances in the complete signal [8], signal processing was applied to separate the voice signal from the excitation signal of the measurement and the response of the vocal tract [16, 17]. A new method relying on multiple chirplet transforms has recently been proposed [18–20] which also investigates the effect of a distant microphone position.

2.4 Miniaturisation

Portable devices have been created in two institutions, the RWTH Aachen University in 2003, and the Singapore University of Technology and Design in 2018 [17, 21]. The device developed at RWTH Aachen can be worn with a holder around the neck and contains two sensors (see Figures 3 and 4). An application of the miniature device is shown in Fig. 5.

Following the successful implementation of the technique of broadband excitation at the speaker/singer's lips and the estimation of vocal tract resonances from its corresponding relative acoustic impedance spectra [22], a modified, low-cost, portable, pocket-sized, and simplified version of

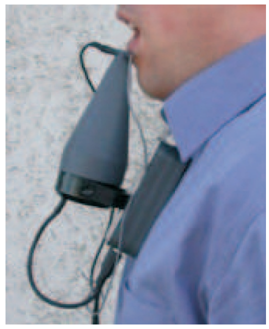


Figure 5. Use of the miniature VTMI device

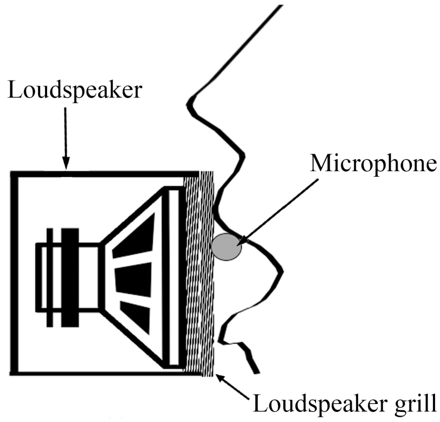


Figure 6. Schematic of the configuration of ACUZ-Lite during measurement (from [17])

this technique named "ACUZ-Lite" was developed [17] for estimating the vocal tract resonances non-invasively and thereby enabling an 'ecological' way of tracking phonation gestures. Whereas the predecessor version (from [22]) required a set of dedicated and high-performance equipment and laboratory conditions to perform the measurements, the ACUZ-Lite, on the other hand, is relatively simple in terms of the hardware and software design.

This compact device (<200 grams) is constructed with a small entry-level loudspeaker and a lavalier microphone, and it is meant for performing 'quick and dirty' field measurements outside the formal laboratory environment. The device is configured in a way that the speaker/singer's lips touch the loudspeaker grill, and the loudspeaker radiates directly into the mouth opening during the measurements as shown in Fig. 6. Similar to the predecessor version, the measurement procedure of the ACUZ-Lite is a two-stage process: a calibration followed by the measurement during phonation. During the calibration process, the speaker's mouth is kept closed, the device is held in the same way shown in Fig. 6, and a broadband noise signal is iteratively modified for this configuration to attain a uniform energy distribution over the appropriate frequency range (typically from 100 Hz to 4000 Hz). Thereafter, the measurement is performed by exciting the vocal tract with the modified broadband noise signal while the speaker phonates a target vowel. Since the loudspeaker drives with considerable power at low amplitude, and also the loudspeaker cone is

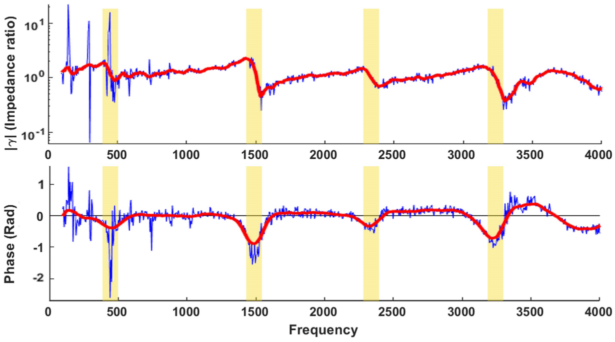


Figure 7. Relative acoustic impedance spectrum measured for the vowel /ə/ using ACUZ-Lite [17]: magnitude (top) and phase (bottom) plots

stiffer than the air, the acoustic velocity of the signal is approximated to be constant for the specified frequency range. Hence a relative acoustic impedance (Gamma) spectrum is evaluated between measured and calibrated reference loads by estimating the ratio of the corresponding pressure signals (assuming $\tilde{v}(t)$ = constant). When plotting the Gamma spectrum, the vocal tract resonances are reflected as 'steep negative slopes' in the magnitude plot and the 'sharp dips' in the phase plot. Figure 7 shows a typical Gamma spectrum obtained for the vowel /ə/ (target word 'herd') that includes a raw (blue) and smoothed (red) version of the measurement. The first four resonance frequencies identified from both magnitude and phase spectra are highlighted.

According to a recent study on the device, tuning the physical configuration of the transducers and the implementation of acoustic foam is shown to improve the acoustic coupling between the device and the vocal tract and develop the sensitivity in low-frequency regions [21]. In order to investigate the optimal physical configuration of transducers, new parameters from the magnitude and phase domains are established in this study for quantitatively assessing the quality and sensitivity of vocal tract resonance detection.

Although the device relies on some approximations and 'quick and dirty' measurements, the resulting relative impedance spectra are comparable in performance with previous studies made under formal laboratory conditions. Despite the absence of acoustic flow information, the device is shown to detect vocal tract resonance frequencies reliably, especially the first resonance frequency, down to 350 Hz and possibly even lower by improving its low-frequency response, and the acoustic coupling with the vocal tract. This is expected to be further enhanced by implementing a more powerful loudspeaker and larger housing, as well as a short inverse horn for acoustic coupling between the device and the vocal tract in case of small mouth openings as in the vowel /u/. Even without further modifications, the device has extended the applicability of the impedance measurement at the mouth to a wide range of vowel articulations and singing styles under 'field conditions'.

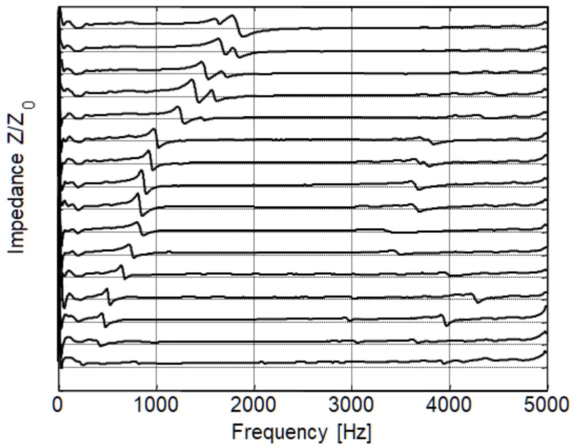


Figure 8. VTMI analysis of a sequence of rising overtones [23]

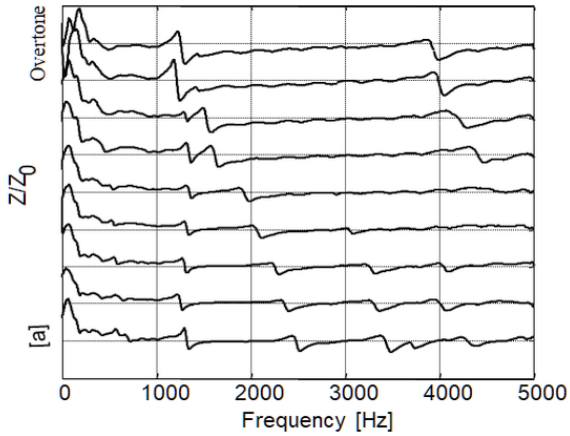


Figure 9. VTMI analysis of a morphed sound from /a:/ to an overtone sound [23]

3. APPLICATIONS

3.1 Analysis of overtone singing

The extraordinarily low bandwidth of the formant that occurs at the frequency of the melody pitch of singers who perform sygyt-style overtone singing, can be explained by focalisation, i.e. concentration of two resonances on one frequency [23]. The result from the impedance analysis of a rising overtone sequence is shown in Fig. 8 and demonstrates the presence of two resonances for some of the higher melody pitches. The impedance measurement of the vocal-tract has been used as a method to demonstrate the morphing of a sustained sound /a:/ to an overtone sound. The formant F_3 lowers its frequency from around 2 kHz to the frequency of F_2 around 1.2 kHz during the morphing process.

The resulting overtone formant has a lower bandwidth and therefore a higher selectivity than each of the formants alone. The transition of a vowel sound /a:/ to an overtone sound shown in Fig.9 demonstrates the focalisation of two resonances that jointly create the strong overtone formant.

3.2 Resonance strategies in operatic singing

An active collaboration started in 2004 between an australian and a french research team, receiving support from a French-Australian PHC funding program for scientific and technological exchanges (FAST, 2010-2011). Resonance-tuning strategies used by opera singers have been studied as a function of pitch, vocal intensity, and sung vowels. Several databases have been compiled:

- in 2004: 22 singers, from amateur to professional level (4 baritones, 8 tenors, 4 altos and 6 sopranos) [LYR2004 database]
- in 2005: 14 professional singers, mostly from the Sydney Opera House (2 baritones, 7 tenors, 5 sopranos) [LYR2005 database]
- in 2008: 12 singers from amateur to professional level (recorded by Maëva Garnier during her post-doctoral stay) [LYR2008 database]
- in 2010: 3 renowned professional singers, mastering yodeling technique

The analyses of these databases have revealed different resonance-tuning strategies in operatic singing. The frequencies of the first two vocal-tract acoustical resonances remained relatively constant throughout the singers' tessitura, as long as the sung note remains lower than the first-resonance frequency [24]. These resonance frequencies can sometimes increase with an increase in fundamental frequency, reflecting a movement of jaw opening as the singer sings higher in pitch. In cases where the first-resonance frequency (f_{R1}) joins the fundamental oscillatory frequency (f_{o}), for instance in the higher part of the tessitura or for closed vowels, a tuning [(f_{R1}):f_{o}] is observed.

This phenomenon of resonance tuning is frequent in sopranos who sing at pitches where resonance and harmonics meet, whatever the training level of the singer. It is also found in other tessituras. Alto singers demonstrate [(f_{R1}):f_{o}] tuning for vowels with a low-frequency first resonance. In the lower part of their pitch range, they may use a [(f_{R1}):$2f_{\mathrm{o}}$] strategy. Baritones and tenors demonstrate [(f_{R1}):$2f_{\mathrm{o}}$] and [(f_{R1}):$3f_{\mathrm{o}}$] tunings in part of their pitch range, or tuning with higher harmonics in low-pitch ranges. Inter-individual differences are great, especially for low-pitch voices.

First-resonance to first-harmonic tuning may be accompanied by an additional frequency tuning to the second resonance [(f_{R2}):$2f_{\mathrm{o}}$] for some singers. To produce their highest-pitch sounds, some of the singers extend their tuning range [(f_{R1}):f_{o}] up to 1300-1500 Hz (E6-F#6). Other singers tune the second resonance near the first harmonic [(f_{R2}):f_{o}] to 2350 Hz (D7).

3.3 Resonance strategies in Bulgarian women's singing

If lyrical singers sometimes present a (f_{R1}) resonance tuning to the second harmonic $2f_{\mathrm{o}}$, the [(f_{R1}):$2f_{\mathrm{o}}$] tuning can be a systematic tuning strategy in other singing styles. For example, Balkan songs, and in particular traditional Bulgarian women's singing, are particularly sonorous, with a predominance of acoustic energy in the voice second harmonic as a key feature [25, 26]. In collaboration with Mara Kiek (ACARMP Sydney Conservatorium of Music, The University of Sydney, Australia), the vocal behavior of an Australian singer who mastered traditional Bulgarian singing techniques was explored. Two distinct vocal qualities were assessed : the teshka (heavy) quality, which corresponds to a very sonorous vocal production, and the leka (light) quality, for which vocal production is softer, approaching in timbre the register of female head voice. The analysis of vowels sung in traditional Bulgarian style has shown the search for a tuning between the first resonance and the closest harmonic, either $2f_{\mathrm{o}}$ or $1f_{\mathrm{o}}$ depending on the vowel, no matter the technique used [27]. This systematic resonance-tuning strategy clearly distinguishes the vocal behavior in Bulgarian singing from the behavior in lyrical singing for the studied female singer. As shown in Figure 10, the resonance frequencies for vowels sung in head voice remain close to the resonance frequencies for the same vowels spoken.

3.4 Analysis of high tenor voices

The impedance measurement has been applied in a larger project for the analysis of potential register transitions in the high range of counter tenor singers. A comparison of the resonances during the performance of a sequence of sustained vowels allowed for grouping of the singers into one group that held the vocal tract configuration unchanged during the whole sequence, and another group that adopted vocal tract resonances to the frequencies of the first three partials. A publication of the work is pending.

4. CONCLUSIONS

The measurement method of vocal tract impedance, invented by Joe Wolfe and John Smith in 1995 has inspired research teams from various countries to further develop and apply it for investigations in speech and singing voice research. Different methods and their applications that had emerged from the technique of vocal tract impedance measurement at the human lips are presented in this review article. This included devices with both pressure and velocity sensors to measure the acoustic impedance, and relatively simple devices that include only a pressure sensor to estimate relative acoustic impedance. Whereas some of the devices are advanced for estimating the voice and resonance qualities for support of research on resonance phenomena in singing, a few have evolved in the direction of simple and portable use in field conditions. 25 years after the initial invention Matlab-based tool with instructions for a self-made device has been published on the UNSW web page of Joe Wolfe [28]. Further improvement of algorithms for the resonance analysis during phonation will enable the use of such devices for singing students outside acoustic laboratories.

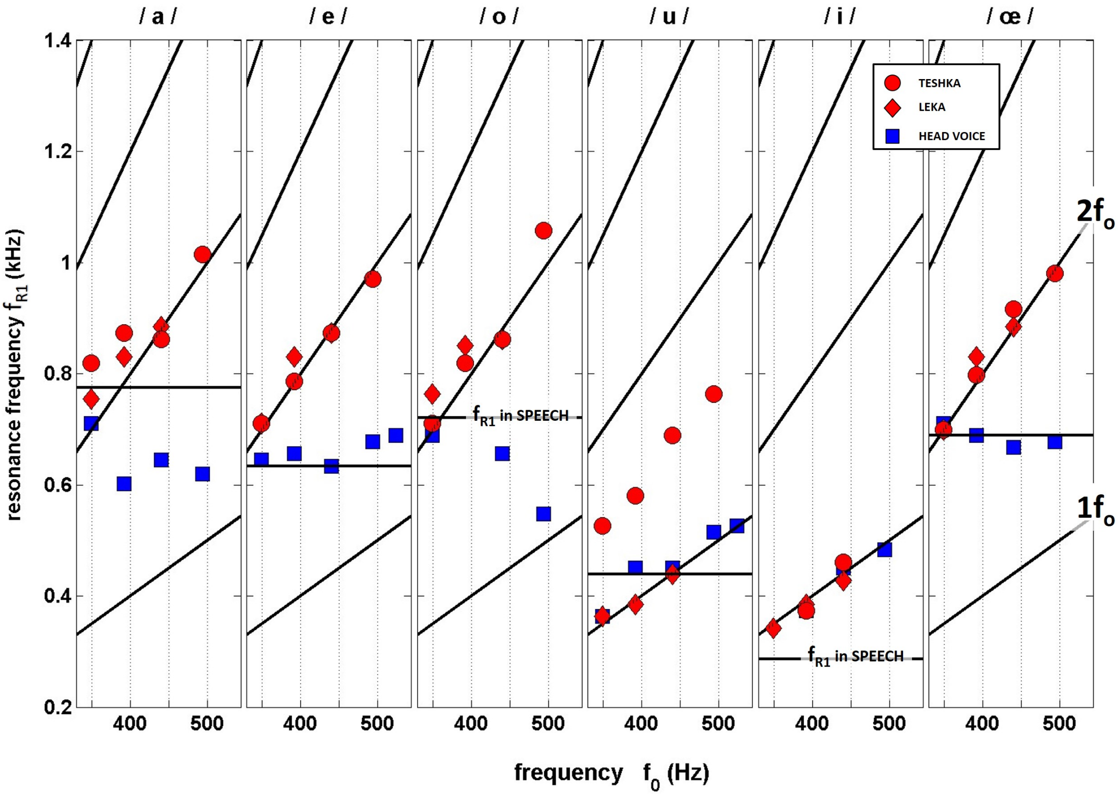


Figure 10. First-resonance frequencies for spoken vowels (horizontal black and sung vowels in bulgarian singing style (red markers) and classical female head voice (blue markers).

Acknowledgments

We acknowledge the contributions of a large number of students, co-workers and colleagues from several countries to the development of the various impedance measurement methods, namely Jer-Ming Chen, Annette Dowd, Julien Epps, Elodie Joliveau, Roman Katzer, Maëva Garnier, Timothée Maison, Luis Roca and Julia Stoffers. Dusty Eddy is thanked for proofreading the manuscript. The reviewers are acknowledged for their constructive comments that helped to improve the manuscript.

5. REFERENCES

[1] J. Wolfe and J. Smith, "A comparison of acoustic impedance of flutes – a preliminary study," in *Proceedings of the International Symposium on Musical Acoustics*. Dourdan, France: Societe Francaise d'Acoustique, 7 1995, pp. 100–106.

[2] A. Dowd, "Real time non-invasive measurements of vocal tract impedance spectra and applications to speech training," Undergraduate thesis, UNSW Sydney, Medical Physics, UNSW Sydney, 1995.

[3] O. Fujimura and J. Lindqvist, "Sweep-tone measurements of vocal-tract characteristics," *The Journal of the Acoustical Society of America*, vol. 49, pp. 541–558, 1971.

[4] F. Ahmadi and I. V. McLoughlin, "Measuring resonances of the vocal tract using frequency sweeps at the lips," in *Proc. 5th International Symposium on Communications, Control and Signal Processing*, 2012, pp. 1–5.

[5] M. Fleischer, A. Mainka, S. Kürbis, and P. Birkholz, "How to precisely measure the volume velocity transfer function of physical vocal tract models by external excitation," *PLOS ONE*, vol. 13, no. 3, p. e0193708, 3 2018.

[6] J. Stoffers, "Akustische Impedanzanalyse der Artikulation," Ph.D. thesis, RWTH Aachen University, 2008. [Online]. Available: https://www.logos-verlag.de/cgi-bin/engbuchmid?isbn=1983&lng=deu&id=

[7] J. Wolfe, J. Smith, J. Tann, and N. Fletcher, "Acoustic impedance spectra of classical and modern flute," *Journal of Sound and Vibration*, vol. 243, no. 1, pp. 127–144, 5 2001.

[8] J. Epps, A. Dowd, J. Smith, and J. Wolfe, "Real time measurements of the vocal tract resonances during speech," in *ESCA. Eurospeech97*, Rhodes, Greece, 1997, pp. 721–724, iSSN 1018-4074.

[9] A. Dowd, J. Smith, and J. Wolfe, "Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time," *Language and Speech*, vol. 41, no. 1, pp. 1–20, 1998. [Online]. Available: https://doi.org/10.1177/002383099804100101

[10] E. Joliveau, J. Smith, and J. Wolfe, "Vocal tract resonances in singing: The soprano voice," *The Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2434–2439, 2004.

[11] R. Vos, H. Daffern, and D. Howard, "Resonance tuning in three girl choristers," *Journal of Voice*, vol. 31, no. 1, pp. 122–e1, 2017. [Online]. Available: https://eprints.whiterose.ac.uk/99829/1/PIIS0892199716000291.pdf

[12] I. R. Titze, R. J. Baken, K. W. Bozeman, S. Granqvist, N. Henrich, C. T. Herbst, D. M. Howard, E. J. Hunter, D. Kaelin, R. D. Kent, J. Kreiman, M. Kob, A. Löfqvist, S. McCoy, D. G. Miller, H. Noé, R. C. Scherer, J. R. Smith, B. H. Story, J. G. Švec, S. Ternström, and J. Wolfe, "Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization," *The Journal of the Acoustical Society of America*, vol. 137, no. 5, pp. 3005–3007, May 2015, publisher: Acoustical Society of America. [Online]. Available: https://asa.scitation.org/doi/10.1121/1.4919349

[13] R. Katzer, "Algorithmen für ein System zur akustischen Artikulationsanalyse," Diploma thesis, RWTH Aachen University, 2003.

[14] J. Stoffers, C. Neuschaefer-Rube, and M. Kob, "Comparison of Vocal Tract Resonance Characteristics Using LPC and Impedance Measurements," *Acta Acustica United with Acustica*, Oct. 2006, publisher: S. Hirzel Verlag. [Online]. Available: https://www.ingentaconnect.com/content/dav/aaua/2006/00000092/00000005/art00004

[15] M. Kob and C. Neuschaefer-Rube, "A method for measurement of the vocal tract impedance at the mouth," *Medical Engineering & Physics*, vol. 24, no. 7, pp. 467–471, Sep. 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1350453302000620

[16] B. Delvaux and D. Howard, "Sinesweep-based method to measure the vocal tract resonances," in *Proceedings of the 9th MAVEBA*. Florence, Italy: Firenze University Press, 2015.

[17] J. Thilakan, B. Balamurali, and J. Chen, "Acuzlite: Ultra-portable real-time estimation of vocal tract resonance," in *Proceedings of the WESPAC 2018*, New Delhi, India, 11 2018.

[18] T. Maison, B. Allain, P. Hoyer, F. Silva, P. Guillemain, and N. Henrich Bernardoni, "Practical guidelines for implementing vocal tract resonances characterization with excitation at the lips," in *Proceedings of the Models and Analysis of Vocal Emissions for Biomedical Applications – MAVEBA: 12^{th} International Workshop*. Florence, Italy: Firenze University Press, 12 2021, p. 31.

[19] T. Maison, F. Silva, N. Henrich Bernardoni, and P. Guillemain, "Séparation et reconstruction d'arêtes par analyse en chirplets d'ordres supérieurs," in *16ème Congrès Français d'Acoustique, CFA2022*, 2022.

[20] T. Maison, F. Silva, P. Guillemain, and N. Henrich Bernardoni, "A new approach to measure vocal tract resonances. application to a modern singing study," in *PEVoC 2022 – 14th Pan-European Voice Conference*, Tallinn, Estonia, Aug. 2022. [Online]. Available: https://hal.science/hal-03763021

[21] J. Thilakan, B. B.T., S. P.M., and J.-M. Chen, "Vocal tract resonance detection at low frequencies: Improving physical and transducer configurations," *Sensors*, vol. 23, no. 2, 2023, article number 939. [Online]. Available: https://www.mdpi.com/1424-8220/23/2/939

[22] J. Epps, J. Smith, and J. Wolfe, "A novel instrument to measure acoustic resonances of the vocal tract during phonation," *Measurement Science and Technology*, vol. 8, no. 10, pp. 1112–1121, 1997.

[23] M. Kob, "Analysis and modelling of overtone singing in the sygyt style," *Applied Acoustics*, vol. 65, no. 12, pp. 1249–1259, Dec. 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003682X04001082

[24] N. Henrich, J. Smith, and J. Wolfe, "Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones," *The Journal of the Acoustical Society of America*, vol. 129, no. 2, pp. 1024–1035, 2011.

[25] G. F. Messner, *Die Schwebungsdiaphonie in Bistrica: Untersuchungen der mehrstimmigen Liedformen eines mittelwestbulgarischen Dorfes*. Tutzing: Hans Schneider, 1980.

[26] P. Boersma and G. Kovacic, "Spectral characteristics of three styles of croatian folk singing," *The Journal of the Acoustical Society of America*, vol. 119, no. 3, pp. 1805–1816, 3 2006.

[27] N. Henrich, M. Kiek, J. Smith, and J. Wolfe, "Resonance strategies used in bulgarian women's singing style: A pilot study," *Logopedics Phoniatrics Vocology*, vol. 32, no. 4, pp. 171–177, 2006.

[28] J. Wolfe, "Vocal resonances and broad band excitation," 2020. [Online]. Available: http://newt.phys.unsw.edu.au/jw/broadband.html

Preliminary acoustic analysis of articulation differences in spoken and sung French language by Greek classical singers

Georgios Papadimitriou
Laboratory of Music Acoustics & Technology (LabMAT), NKUA
geo2pap@yahoo.gr

Areti Andreopoulou
Laboratory of Music Acoustics & Technology (LabMAT), NKUA
aandreo@music.uoa.gr

Anastasia Georgaki
Laboratory of Music Acoustics & Technology (LabMAT), NKUA
georgaki@music.uoa.gr

ABSTRACT

Singing in a foreign language can pose an important challenge to a singer. Vowel sounds that do not exist in one's native tongue are usually the most frequent reason for sounding foreign, while the erroneous production of such phonemes might lead to crucial intelligibility issues. This preliminary study examines the degree to which the substantial knowledge of a foreign language (French) can assist a Greek-speaking classical singer in performing authentically in it. To this end, 16 male native Greek-speaking classical singers recorded excerpts from the standard French-language repertoire of their voice type. Participants provided both spoken and sung audio samples of their voice. The Formant analysis focused on 4 special cases of difficulty of the French language: vowels /ɑ̃/, /ə/, /e/, and /y/ - all foreign to the Greek language. The study revealed the special cases in which there exist pronunciation differences between speakers and non-speakers of the foreign language, as well as articulation differences between speaking and singing voices.

1. INTRODUCTION

Classical singing demands a specific placement of the voice (*singer's formant*), a condition which usually compromises the precise articulation of the sung words [1]. This issue is exacerbated in the (very frequent) case when a singer is required to sing in a foreign language with significantly different articulation from their mother tongue which they haven't fully mastered phonologically and semantically. As a result, enunciation deviations in the syllables between speaking and singing are noticed, not only in the foreign language, but also in the mother tongue of the singer, which can lead to serious intelligibility issues of the sung musical text. All in all, it has long been acknowledged that singing in a foreign language poses a crucial challenge to the interpreter [2].

Lyric diction in a foreign language is an interdisciplinary field of study. Among others, it consists of prosody study, music performance, classical singing teaching, and foreign language learning. Due to this complexity of the issue, there are only few contemporary studies which have substantially dealt with the problem of evaluating lyric diction. On the same research axis, a few researchers have tried to enlighten the issue (e.g. Parpinelli et al. for Brazilian Portuguese [3]), shedding light to issues such as: "How comprehensible is the pronunciation of foreign singers", "What are the main articulation problems that interfere with intelligibility in singing", "How can these problems be overcome in lyric diction teaching". Rosenau [4], on the other hand, conducted an analysis of phonetic differences between German singing and speaking voices, while Cornaz et al. [5] provided one of the most important studies in the field regarding the development of an original learning, teaching and evaluation method supported by speech therapists, linguists, and singers. This last study offers valuable insight into which techniques/practices could be employed by a student of classical singing in order to achieve lyric diction mastery in a foreign language quickly and effectively.

French opera is one of Europe's most important operatic traditions. Offenbach's operettas, Gounod's *Faust*, and Bizet's *Carmen* all debuted in the second half of the 19th century and are now considered some of the world's most famous operatic works [6]. At the same period, the French art song (mélodie) also arose in the fine works of i.a. Faure, Debussy and Ravel. It is safe to admit that today French is an essential language of the classical vocal anthology. In terms of articulation, and in contrast to the Greek language, French is the most complex among the standard operatic languages (Italian, German, and English) due to its numerous vowel sounds and extensive consonant differentiations [7]. Considering all the above, French is regarded as the most interesting language of study for this research.

The primary aim of the present preliminary study is to investigate the degree to which the substantial knowledge of the French (foreign) language − which can be achieved through education, everyday practice in a French-speaking environment, and professional usage − helps a Greek-speaking classical singer to deal effectively with issues of correct pronunciation in a French musical text, so as to manage to sing in a more authentic way in it throughout the range of his/her voice. To that end, this study has focused on 4 cases of vowel sounds that put extra strain on Greek singers: /ɑ̃/, /y/, /ə/, /e/ (IPA symbols), i.e., phonemes that are foreign to the official Greek language.

2. METHODOLOGY

2.1 Subjects & Material

16 male Greek-speaking professional classical singers were assigned to two groups depending on the information they provided in regard to their command of the French language. Thus, group 1 consisted of 8 non-native French speakers and group 2 of 8 non-speakers of French. All subjects belong to the low male voice type ("Fach" of basses, baritones, and bass-baritones).

In an attempt to determine how important it is for classical singers to study the pronunciation, articulation, and enunciation of the French language in order to fully convey the meaning of the sung words and ensure a clearer understanding by the audience, participants were asked to sing excerpts taken from the standard repertoire of their voice type. This way, the vowels – to be subsequently isolated and examined – would be obtained from a typical and expected syllabic context, just as they exist in the classical music vocal anthology. This approach was considered more representative of “a real-life situation” than recording and examining vowels in isolation. This kind of experiment, however, added two challenges in the analysis: i) in every syllable, the vowels examined were preceded by a different consonant, and ii) every examined syllable was in a different tonal pitch.

The study was not a foreign-language knowledge test. Participants were asked to sing excerpts which had been communicated to them well in advance (10 to 15 days beforehand), allowing them enough time to prepare for the recording process in the best possible – to their discretion – way. No participants have reported any voice or hearing problem, while their ages range from 26 to 54. All of them had at least two years of professional experience as classical singers and a similar minimum educational background.

The following 19th-century composition excerpts were used:

1. La nuit sur le grand mystère, entr’ ouvre ses écrins bleus − *Nocturne* (G. Fauré), measures 1-5
2. On voit ses ombres dormantes, S'éclairer, à tous moments − *Nocturne* (G. Fauré), measures 15-19
3. N’a pour charme et pour clarté, Qu’une fleur, et qu’une étoile − *Nocturne* (G. Fauré), measures 34-37
4. Qu’un mot, peut rendre malheureux, hélas − Quand la flamme de l’amour [Ralph], *La jolie fille de Perth* (G. Bizet), measures 29-30
5. Où brille l’ardent métal − Le veau d’or [Méphistophélès], *Faust* (Ch. Gounod), measures 32-35
6. A l’amant qui vous implore, pourquoi refuser, un si doux baiser − Vous qui faites l’endormie [Méphistophélès], *Faust* (Ch. Gounod), measures 33-36 (from the beginning of the 1st verse).

These excerpts were carefully selected, so as to lie within the middle area of the low male voice [8], and specifically from E♭3 to B3 (F0 = 155.56 Hz - 246.94 Hz). It was a deliberate choice not to use excerpts with extreme emotional content or excerpts that included extremely high or extremely low notes, as in these cases a vowel disintegration and a loss of its clarity is observed anyways [9].

2.2 Recordings & measurements

Participants were recorded in an acoustically-treated space, in the facilities of the *Laboratory of Music Acoustics and Technology (LabMAT), NKUA,* providing both spoken and sung samples of their voice. In particular, they first recited and then sang acapella the poetic text (libretto) of each excerpt (aria or mélodie), using the M1 laryngeal vibratory mechanism of their voice.

Recordings were conducted using a Neumann-KM 184 condenser microphone, making use of a ‘HV-3 Millenia Media’ preamplifier and a RME converter (model: ADI-8DS). The microphone was placed at a distance of 0.5 m and at an angle of 20° from the singer’s lips. Participants were advised to stand as still as possible during the recording process. Samples were recorded at a sampling rate of 48 kHz and a bit rate of 24 bits using CUBASE 11.

3. ANALYSIS AND RESULTS

The recordings were analyzed using PRAAT. The vowel sounds to be examined were isolated in each syllable where a different consonant sound was preceded (as they appeared in the musical excerpts), their spectrograms were printed out, and the first two Formants of each vowel sound were recorded, along with the fundamental frequency in which they were produced (spoken or sung).

The Formant values for those 4 cases under study (/ã/, /y/, /ə/, /e/) were analyzed in SPSS, separately for each case. The variable ‘Speaker Group’ has two values: “non-native speaker”, and “non-speaker” of the foreign language (French). The variable ‘Preceding Consonant’ which corresponds to the different consonants or consonant combinations that preceded each vowel sound under examination differ and are described in detail in the respective subparagraphs that follow. In each case, a two-way ANOVA was conducted separately for each Formant, both for speaking and for singing, in order to examine the interaction of ‘Speaker Group’ and ‘Preceding Consonant’ in the values of F1 and F2. In this way, it would be revealed whether the mean scored values of F1 and F2 differ significantly in relation to the linguistic background of the speaker/singer and the articulation particularities of the preceding consonants. According to Sundberg [10], “Formant frequencies depend on articulation”, while the first and second Formant are used to depict the clarity of the vowel.

Figure 3, which represents the French vowel chart by Collins & Mees [11], as adapted by G. Papadimitriou based on data by the *International Phonetic Association*, will be used as a frequent reference in this analysis regarding the expected values of F1 and F2, and specifically their “positions” on a two-dimensional chart [12], namely the “two major articulatory dimensions (mandibular and lingual, respectively) in vowel production” [13].

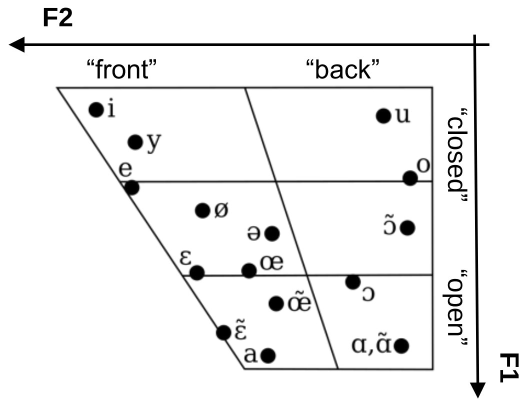


Figure 3. French vowel chart (Collins & Mees) [11], adapted by G. Papadimitriou

3.1 Vowel sound /ã/

The 'preceding consonant' combinations for the /ã/ vowel include: the consonant complex /gr/, the "bare" vowel /-/, the consonant complex /rm/, the consonant /m/, and the consonant /r/.

The results indicated a statistically significant difference in the mean values of the second Formant (F2) between the two different speaker-groups when singing ($F(1, 70) = 9.354$, $p = 0.003$). Pairwise comparisons of the means revealed that 'Non-native Speakers' recorded a lower mean value for F2 ($M = 939$, $SD = 100$) in comparison to 'Non-Speakers' ($M = 1038$, $SD = 182.5$).

Significant difference in the F2 mean values appeared also between the two speaker-groups ($F(1, 70) = 11.446$, $p = 0.001$) and among the five different consonant combinations ($F(4, 70) = 3.215$, $p = 0.018$) in the speaking condition. Pairwise comparisons of the mean values revealed that 'Non-native Speakers' recorded a lower mean F2 value ($M = 923.75$, $SD = 141.85$) in comparison to 'Non-Speakers' ($M = 1034.83$, $SD = 164.261$). In addition, when reciting the poetic text the preceding consonant /m/ led to a significantly lower scores in F2 in both 'speaker-groups' ($M < 900$). In the rest of the examined cases – although not 'statistically significant' – it is interesting to mention that the preceding consonant /m/ tends to decrease the values of both Formants, in both 'speaker-groups', whether speaking or singing.

There were no other statistically significant differences found between the other conditions ($p > 0.05$; see Table 3.1). Mean values and 95% confidence intervals (CI) of F1 and F2 for /ã/ are illustrated in Figures 3.1.1 and 3.1.2 respectively.

/ã/	F1		F2	
Speaking	F	p-value	F	p-value
Speaker group	0.536	0.467	11.446	0.001
Consonant	1.559	0.195	3.215	0.018
Variable Interaction	0.599	0.665	0.588	0.672
/ã/	F1		F2	
Singing	F	p-value	F	p-value
Speaker group	1.047	0.303	9.354	0.003
Consonant	1.388	0.247	1.89	0.122
Variable Interaction	0.363	0.834	0.89	0.475

Table 3.1. Concise results of two-way Anova in Formants of /ã/

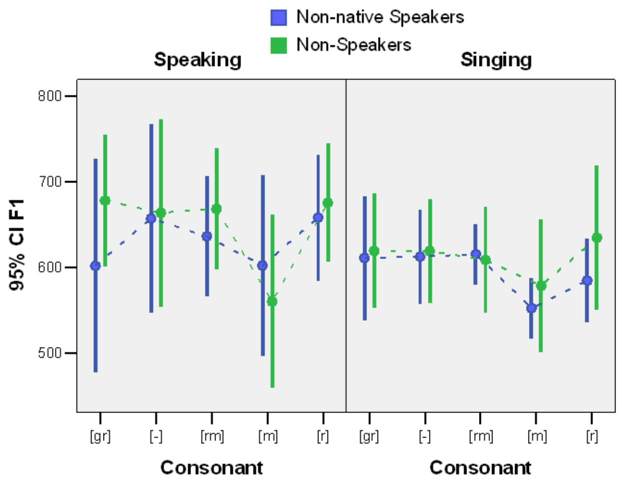


Figure 3.1.1. Mean values and 95% confidence intervals of F1 (Hz) for /ã/ for the two 'speaker-groups', in each 'preceding consonant' combination examined.

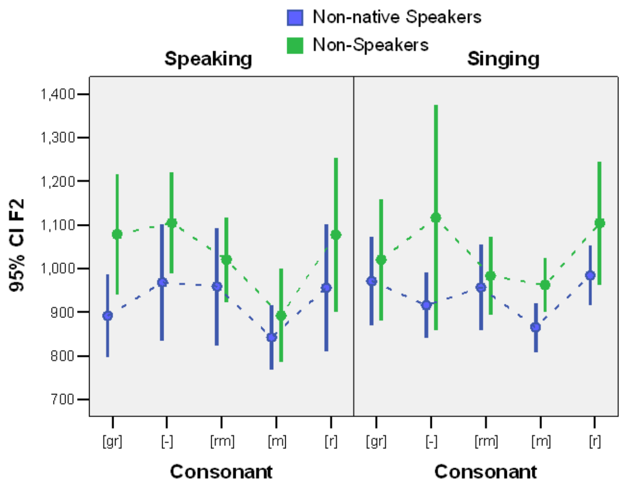


Figure 3.1.2. Mean values and 95% confidence intervals of F2 (Hz) for /ã/ for the two 'speaker-groups', in each 'preceding consonant' combination examined.

Given the results above and looking at Figure 3 it can be observed that the increased F2 values for vowel sound /ã/ in the 'non-speakers' group tend to approach those expected for vowel /a/. This applies both to speaking and singing. It can be assumed that 'non-speakers' tend to speak and sing these vowels "more forward" [12], as if they were singing a phoneme closer to /a/, i.e., a vowel that exists in their mother tongue.

The overall F2 values seem to range in the same area in both 'Speaking' and 'Singing' mode. The overall F1 values, instead, are slightly increased when speaking ($M = 640.08$, $SD = 110.718$) in comparison to singing ($M = 603.50$, $SD = 72.418$) giving us evidence for articulation modification concerning either a greater jaw opening when speaking /ã/ or a rounding of the lips when singing /ã/[10].

3.2 Vowel sound /ə/

The 'preceding consonant' combinations for the /ə/ vowel include the consonants: /r/, /n/, /l/, /t/, and the semi-vowel /j/.

In singing, the results indicated a statistically significant difference in the mean values of the first Formant (F1) among the five different ‘preceding consonants’ ($F(4, 70) = 18.339$, $p = 0.000$). Pairwise comparisons of the means using the Scheffe criterion revealed higher F1 values for consonant /n/, and semi-vowel /j/ ($M > 540$). The results also indicated a statistically significant difference in the mean values of the second Formant (F2) between the two different ‘speaker-groups’ ($F(1, 70) = 10.628$, $p = 0.002$). Further pairwise comparisons showed that ‘Non-native Speakers’ recorded a lower mean value for F2 ($M = 1241$, $SD = 113.4$) in comparison to ‘Non-Speakers’ ($M = 1328$, $SD = 126.8$).

On the contrary, in speaking, significant differences in the F2 mean values appeared both between the two speaker-groups ($F(1, 70) = 10.325$, $p = 0.002$) and among the five different consonant combinations ($F(4, 70) = 8.477$, $p = 0.000$). Pairwise comparisons of the means revealed that ‘Non-native Speakers’ recorded a lower mean value for F2 ($M = 1362.75$, $SD = 195.202$) in comparison to ‘Non-Speakers’ ($M = 1466.40$, $SD = 168.515$). In addition, when reciting the poetic text the preceding consonants /r/ and /t/ led to significantly lower scores in F2 in both ‘speaker-groups’ ($M < 1400$).

There were no other statistically significant differences found between the other conditions ($p > 0.05$; see Table 3.2). Mean values and 95% confidence intervals (CI) of F1 and F2 for /ə/ are illustrated in Figures 3.2.1 and 3.2.2 respectively.

/ə/	F1		F2	
Speaking	F	p-value	F	p-value
Speaker group	2.525	0.117	10.325	0.002
Consonant	2.134	0.086	8.477	0.000
Variable Interaction	0.592	0.67	0.24	0.914
/ə/	F1		F2	
Singing	F	p-value	F	p-value
Speaker group	2.072	0.155	10.628	0.002
Consonant	18.339	0.000	2.02	0.101
Variable Interaction	0.228	0.922	0.402	0.807

Table 3.2. Concise results of two-way Anova in Formants of /ə/

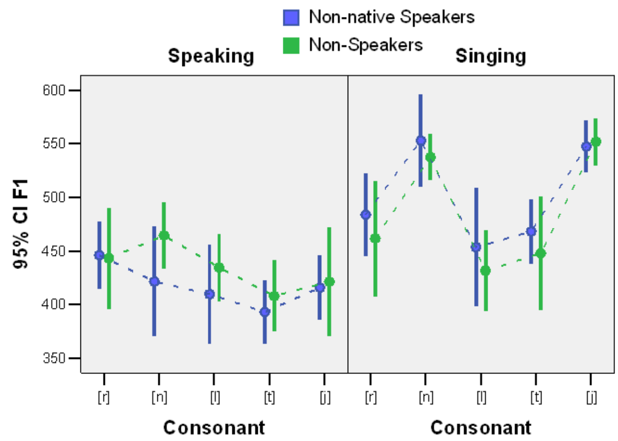


Figure 3.2.1. Mean values and 95% confidence intervals of F1 (Hz) for /ə/ for the two ‘speaker-groups’ in each ‘preceding consonant’ combination examined.

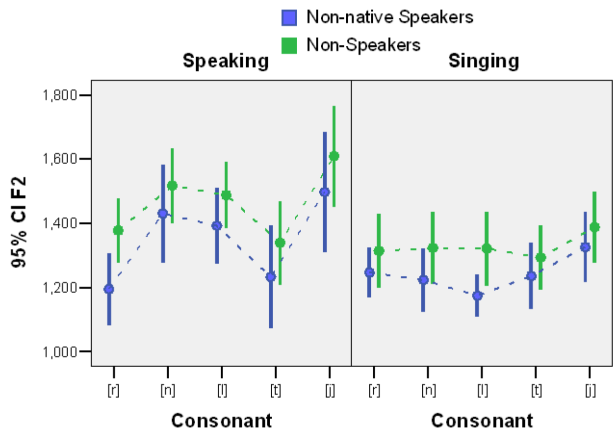


Figure 3.2.2. Mean values and 95% confidence intervals of F2 (Hz) for /ə/ for the two ‘speaker-groups’ in each ‘preceding consonant’ combination examined.

Given the results above and looking at Figure 3 it is observed that the increased F2 values for vowel sound /ə/ in the ‘non-speakers’ group would probably lie closer to those expected for vowel /ɛ/. This applies both to speaking and singing. It can be assumed that ‘non-speakers’ tend to speak and sing /ə/ “more forward” [8], as if they were singing a phoneme closer to /ɛ/, i.e., a vowel that exists in their mother tongue.

The higher F1 values in the sung syllables /nə/ and /jə/ can be explained by the fact that these syllables were sung on a tonal pitch of B♭3 (F0=233.09 Hz). This is the area of *primo passaggio* of the low male voice [14], an area where the singer starts to open his mouth wider by lowering his mandible; this inevitably leads to higher F1 values [12]. Hence, it is deduced that it is the high pitch and not the preceding consonant that caused high F1 values in singing /ə/.

Overall, the values of F1 in speaking ($M = 427.45$, $SD = 47.312$) seem to range lower than the F1 values in singing ($M = 493.64$, $SD = 63.960$). The opposite is observed for the F2 values; the ones in speaking ($M = 1417.30$, $SD = 187.783$) range higher than those in singing ($M = 1284.85$, $SD = 127.241$), information that hints at the possibility of articulation differences between speaking and singing /ə/. Considering all the above, the vowel sound /ə/

is probably spoken with rounder lips and more ‘forward’ than sung.

3.3 Vowel sound /e/

The ‘preceding consonant’ combinations for the /e/ vowel include the consonants: /s/, /r/, /m/, /z/, the consonant cluster /rt/, and the bare vowel /-/.

In singing, the results indicated a statistically significant difference in the mean values of F1 among the six different ‘preceding consonants’ (F(5, 84) = 11.285, p = 0.000). Pairwise comparisons of the means using the Scheffe criterion revealed very low F1 values for consonant cluster /rt/ (M < 431), and very high F1 values for consonant /m/ (M > 520).

On the contrary, in speaking, significant differences in the F1 mean values appear both between the two speaker-groups (F(1, 84) = 6.792, p = 0.011) and among the six different consonant combinations (F(5, 84) = 4.755, p = 0.001). Pairwise comparisons of the means reveal that ‘Non-native Speakers’ scored a lower mean value for F1 (M = 410.48, SD = 61.406) in comparison to ‘Non-Speakers’ (M = 439.17, SD = 54.873). Furthermore, just as in the “singing situation”, the preceding consonant cluster /rt/ led to very low F1 values (M < 412), and the consonant /m/ to very high F1 values (M > 472). In regard to the F2 values, significant difference appears only between the two speaker-groups (F(1, 84) = 19.955, p = 0.000). Pairwise comparisons of the means revealed that ‘Non-native Speakers’ scored a higher mean value for F2 (M = 1837.75, SD = 131.584) in comparison to ‘Non-Speakers’ (M = 1718.69, SD = 130.152).

There were no other statistically significant differences found between the other conditions (p > 0.05; see Table 3.3). Mean values and 95% confidence intervals (CI) of F1 and F2 for /e/ are illustrated in Figures 3.3.1 and 3.3.2 respectively.

/e/	F1		F2	
Speaking	F	p-value	F	p-value
Speaker group	6.792	0.011	19.955	0.000
Consonant	4.755	0.001	1.470	0.208
Variable Interaction	0.366	0.870	0.616	0.688
/e/	F1		F2	
Singing	F	p-value	F	p-value
Speaker group	2.132	0.148	3.138	0.08
Consonant	11.285	0.000	1.141	0.345
Variable Interaction	0.622	0.683	0.226	0.95

Table 3.3. Concise results of two-way Anova in Formants of /e/

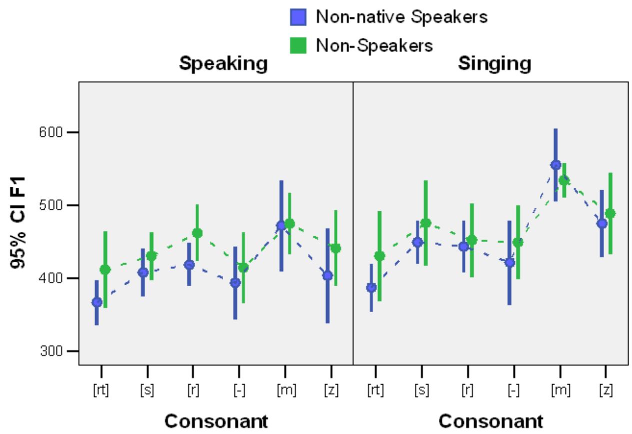


Figure 3.3.1. Mean values and 95% confidence intervals of F1 (Hz) for /e/ for the two ‘speaker-groups’ in each ‘preceding consonant’ combination examined.

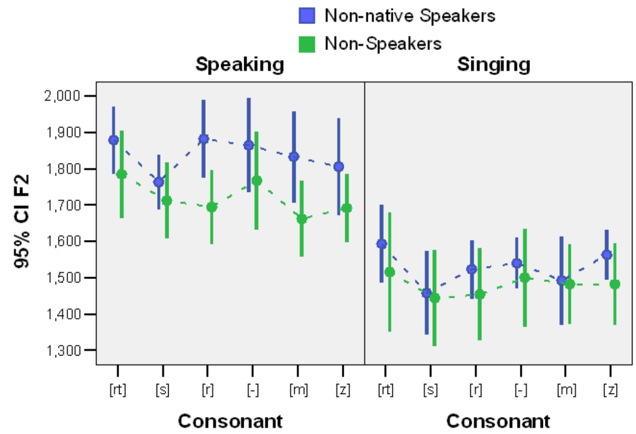


Figure 3.3.2. Mean values and 95% confidence intervals of F2 (Hz) for /e/ for the two ‘speaker-groups’ in each ‘preceding consonant’ combination examined.

All in all, there was not enough evidence to assert that singing /e/ varies significantly between the two speaker-groups. However, when speaking it, the higher F1 values in tandem with the lower F2 values scored by ‘non-speakers’ indicate that they probably articulate /e/ closer to /ɛ/, based on examination of Figure 3.

The higher F1 values which consonant /m/ provoked in singing can be explained by the fact that the syllable which contained it was sung on a tonal pitch of B3 (F0=246.94 Hz). As explained in subparagraph 3.2., this is the area of *primo passaggio* of the low male voice [14], which inevitably leads to higher F1 values [12]. Hence, it is deduced that it is the high pitch and not the preceding consonant that caused high F1 values in singing /e/.

Overall, the F1 values in speaking (M = 424.82, SD = 59.692) seem to range relatively close to the F1 values in singing (M = 463.51, SD = 68.990). However, the F2 values in speaking (M = 1778.22, SD = 143.276) range significantly higher than those in singing (M = 1503.89, SD = 134.032), giving us some clue for articulatory differentiation between speaking and singing. According to Clermont [13] and Florig [15], lower F2 values are expected for a ‘front’ vowel when sung.

3.4 Vowel sound /y/

The ‘preceding consonant’ combinations for the /y/ vowel include the consonants: /s/, /k/, /f/.

In singing, the results indicated a statistically significant difference in the mean values of the first Formant (F1) among the three different ‘preceding consonants’ ($F(2, 42) = 4.929$, $p = 0.012$). Pairwise comparisons of the means using the Scheffe criterion revealed that the consonant /s/ led to significantly lower values for F1 (*M* = 321, *SD* = 17,8) than the consonant /f/ (*M* = 377, *SD* = 48,5).

A similar situation was noticed in speaking as well. A statistically significant difference in the F1 mean values was revealed among the different ‘preceding consonants’ ($F(2, 42) = 12.622$, $p = 0.000$). Pairwise comparisons of the means using the Scheffe criterion revealed that the consonant /s/ led to significantly higher values for F1 (*M* = 315, *SD* = 43.4) than the consonant /k/ (*M* = 258, *SD* = 25.7).

There were no other statistically significant differences found between the other conditions ($p > 0.05$; see Table 3.4). Mean values and 95% confidence intervals (CI) of F1 and F2 for /y/ are illustrated in Figures 3.4.1 and 3.4.2 respectively.

/y/	F1		F2	
Speaking	F	p-value	F	p-value
Speaker group	2.355	0.132	2.032	0.161
Consonant	12.622	0.000	3.030	0.059
Variable Interaction	0.492	0.615	1.023	0.368
/y/	F1		F2	
Singing	F	p-value	F	p-value
Speaker group	0.209	0.65	0.132	0.718
Consonant	4.929	0.012	1.45	0.246
Variable Interaction	0.137	0.871	0.076	0.927

Table 3.4. Concise results of two-way Anova in Formants of /y/

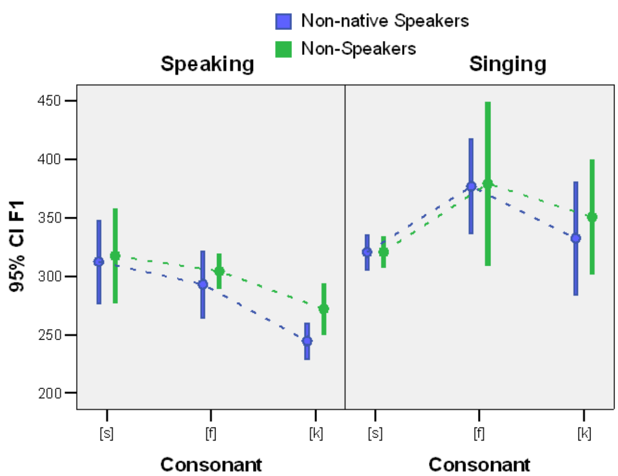


Figure 3.4.1. Mean values and 95% confidence intervals of F1 (Hz) for /y/ for the two ‘speaker-groups’ in each ‘preceding consonant’ combination examined.

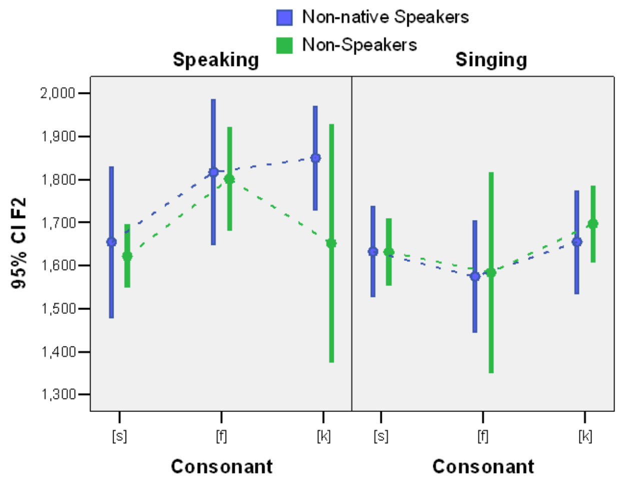


Figure 3.4.2. Mean values and 95% confidence intervals of F2 (Hz) for /y/ for the two ‘speaker-groups’ in each ‘preceding consonant’ combination examined.

There was not enough evidence to assert that the articulation of /y/ varies significantly between the two ‘speaker-groups’, whether speaking or singing it.

The higher F1 values scored in the sung syllable /fy-/ can be explained by the fact that it was sung on a tonal pitch of A3 (F0=220 Hz). Once again, we are dealing with the *primo passaggio* area of the low male voice [14], which unavoidably leads to higher F1 values [12]. Therefore, the high F1 values in sung syllable /fy-/ were caused by the high pitch and not by the preceding consonant /f/.

Overall, the F1 values in speaking (*M* = 290,65, *SD* = 40,228) seem to range lower that the F1 values in singing (*M* = 346,75, *SD* = 55,269); this hints at the possibility that the participants tended to speak /y/ closer to /i/ (based on examination of Figure 3). Yet, the F2 values in speaking (*M* = 1732,69, *SD* = 210,184) range higher than those in singing (*M* = 1628,85, *SD* = 157,543), which can be expected for a ‘front’ vowel [13, 15]. Considering all the above, there is enough evidence for articulatory differentiation between speaking and singing /y/.

3.5 Superimposed vowel charts

In a similar way to the “articulatory vowel chart” (eg. Figure 3), the frequencies of the 1st and 2nd Formant (F1 and F2) can be graphed on an “acoustic vowel chart”. F1 is represented on the vertical axis and F2 on the horizontal axis with the direction of increasing frequency reversed in order to emphasize the relationship to the articulation, as well as to simulate the ‘horizontal’ and ‘vertical axis’ — terms used empirically when teaching singing — of the human voice apparatus.

The four vowel cases examined in this paper form a unique quasi-trapezium in every condition examined, i.e., ‘speaker’ or ‘non-speaker’, ‘singing’ or ‘speaking’. Figures 3.5.1 and 3.5.2 show the different areas that these trapezia occupy when participants of the two speaker-groups speak and sing respectively. It is easily noticed that the ‘singing’ trapezia appear somewhat “condensed” and retracted to the right for both ‘non-native speakers’ and ‘non-speakers’ of French. ‘Front’ vowels (y, e) are the main “culprit” for this retraction.

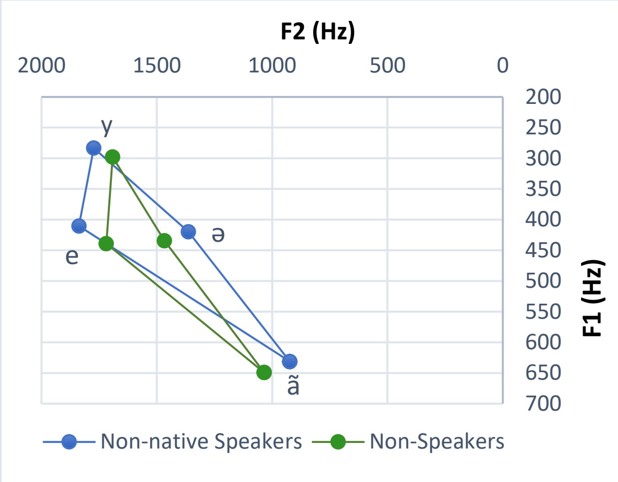


Figure 3.5.1. Vowel Chart for **Speaking** of the two ‘speaker-groups’

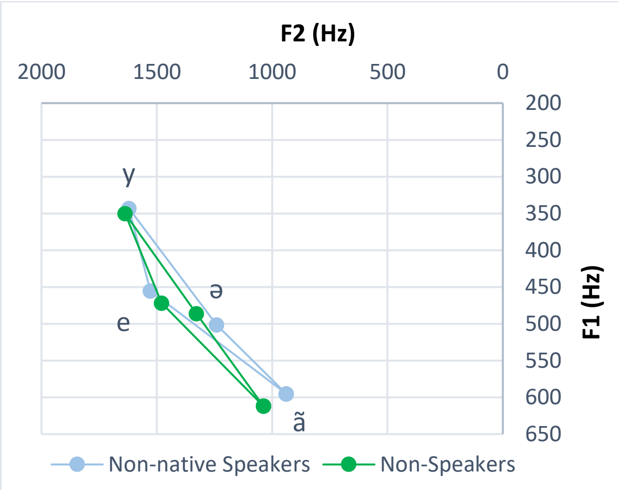


Figure 3.5.2. Vowel Chart for **Singing** of the two ‘speaker-groups’

Figures 3.5.3 and 3.5.4 show the different shapes these trapezia respectively take when examining ‘Non-native Speakers’ and ‘Non-Speakers’ separately. What is noticeable in both cases is that the ‘singing trapezium’ has almost assumed the shape of a ‘quasi-rhombus’ which is virtually included within the ‘speaking trapezium’. Nasal a (ã) is the only vowel that seems to “disobey” this “rule” for Non-native Speakers’, while schwa (ə) also manages to “escape” the ‘speaking trapezium’ in the case of ‘Non-Speakers’.

Last, Figure 3.5.5 provides a general image of the speaking and singing trapezia / rhombi, regardless of the French-language knowledge background of the participants. As in the cases separately examined, the ‘singing polygon’ seems to be almost contained in the ‘speaking polygon’, with the exception of /ã/.

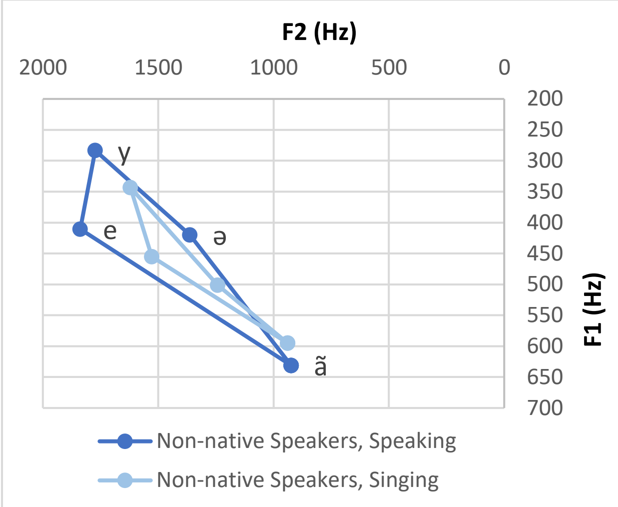


Figure 3.5.3. Vowel Chart for ‘Non-native Speakers’: Speaking vs. Singing

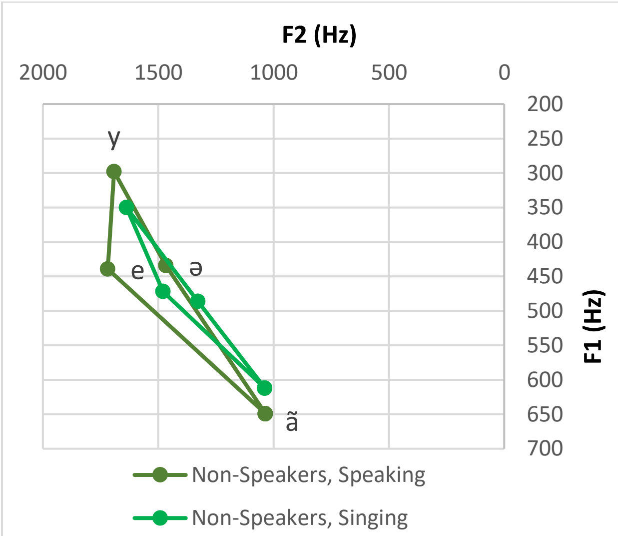


Figure 3.5.4. Vowel Chart for ‘Non-Speakers’: Speaking vs. Singing

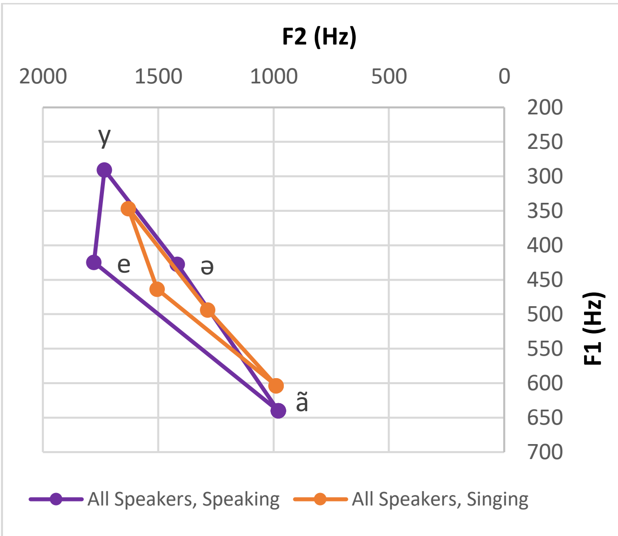


Figure 3.5.5. Vowel Chart for Speaking vs. Singing for all speakers

4. DISCUSSION

The findings of the present study reinforce the authors' strong belief that developing a comprehensive knowledge of phonetics and phonology of French as well as an ability to correctly configure the oral tract to produce the speech sounds of the different vowels and consonants is crucial for the professional classical singer in order to ensure an accurate representation and performance of the French language.

The official Greek educational system is very inefficient at training students of classical singing appropriately so as to get them ready for the demanding operatic world. In most Western European Universities and Academies of music students undertake the study of the basic structures of French in order to develop their ability to fully appreciate the operatic text. The study of French is also often applied to specific roles that students may be undertaking in operas being performed during the semester or to other vocal repertoire.

While this educational approach has been proven very successful throughout the years for every Greek student that has studied abroad, additional research is required in order to treat the whole matter from the acoustics perspective more thoroughly.

5. CONCLUSIONS

The analysis of the four vowel sounds revealed various differences between the two speaker-groups, between speaking and singing mode, and also provided us with valuable clues about when the preceding consonant plays an important role in the succeeding vowel articulation.

Non-speakers of French tended to sing vowel sounds /ɑ̃/ and /ə/ closer to /a/ and /ɛ/. In a similar way, in speaking, they articulated /ɑ̃/, /ə/, and /e/ closer to /a/, /ɛ/, and /ɛ/ respectively.

The preceding consonant seemed to interact in one way or another with all of the vowel cases examined in speaking. However, this was not the case in singing, where tonal pitch was deemed as the primary interaction factor in articulation.

Regarding the differences between speaking and singing, the participants seemed to be employing different articulatory mechanisms in most cases between speaking and singing. The results vary for each vowel sound examined and, in general, have to do with a greater jaw opening when speaking /ɑ̃/, a rounding of the lips when speaking /ə/, and bringing the "closed" spoken vowels /e/ and /y/ more "forward". As a general conclusion, vowels seemed to "lie" closer to each other when singing in comparison to speaking, suggesting that classical singing requires greater homogeneity in the articulation of the phonemes than speaking.

Apart from the 4 phonemes that were examined in the present paper, future studies need to examine other phonemes and transition consonants in different tonal pitches, so as to determine the key factors that prevent foreigners from singing authentically in French. The researchers' ultimate goal is to develop a pedagogical approach of teaching lyric diction that puts extra emphasis on the problematic areas of a Greek-speaking classical singer.

6. REFERENCES

[1] J. Sundberg, C. Romedahl, "Text intelligibility and the singer's formant—a relationship?." Journal of Voice, 23.5, 2009, pp 539-545

[2] G. P. Orwen, The Singer and Foreign Languages. The Modern Language Journal, 1960, 44.1: pp. 17-20.

[3] D. Parpinelli, W. Cardoso, "Singing in a Foreign Language: Can the Fat Lady Pronounce her Phonemes?", Proceedings of the International Symposium on the Acquisition of Second Language Speech, 2014

[4] S. Rosenau, "An analysis of phonetic differences between German singing and speaking voices." In: Proc. 14th Int. Congress of Phonetic Sciences. 1999

[5] S. Cornaz, N. Henrich, N. Vallée, «L'apport d'exercices en voix chantée pour la correction phonétique en langue étrangère : le cas du français langue étrangère appliqué à des apprenants italiens d'âge adulte», Cahiers de l'APLIUT, Vol. XXIX N° 2, 2010, pp. 103-119

[6] L. Orrey, R. Milnes, Opera: A Concise History. World of Art. Thames & Hudson. 1987, ISBN 0-500-20217-6.

[7] D. Adams, A handbook of diction for singers: Italian, German, French. Oxford University Press, 2008

[8] D. G. Miller, H. K. Schutte, "Toward a Definition of Male 'Head' Register, Passaggio, and 'Cover' in Western Operatic Singing". Folia Phoniatrica Logopaedica, 46, 1994, pp 157-170

[9] H. Hollien, A. P. Mendes-Schwartz, K. Nielsen, "Perceptual confusions of high-pitched sung vowels", Journal of Voice, Volume 14, Issue 2, 2000, pp 287-298

[10] J. Sundberg, The science of singing voice. Northern Illinois University Press, 1987

[11] B. Collins, I. M. Mees, Practical Phonetics and Phonology: A Resource Book for Students (3rd ed.), Routledge, 2013, pp. 225–226

[12] J. Sundberg, "The acoustics of the singing voice." Scientific American 236.3, 1977, pp. 82-91

[13] F. Clermont, "Systemic comparison of spoken and sung vowels in formant-frequency space", Proceedings of the 9th Australian International Conference on Speech Science & Technology (SST), 2002

[14] R. Miller, Securing baritone, bass-baritone, and bass voices. pp. 8-9, Oxford University Press, 2008

[15] E. Florig, An acoustic study of French vowels in speech and singing voice. In: Proceedings of the International Conference of Phonetic Sciences (ICPhS'95), Stockholm. 1995, p. 210-213.

A PILOT STUDY OF VOCAL VIBRATO INCORPORATING NONLINEAR TIME SERIES ANALYSIS

Gerardo Acosta Martínez & Helena Daffern
AudioLab, University of York, Heslington, York, United Kingdom
gerardo.acosta@york.ac.uk
helena.daffern@york.ac.uk

ABSTRACT

Research on vocal vibrato suggests that regularity and periodicity can affect its auditory perception and pleasantness. Currently, vibrato regularity has been mostly measured by analyses such as jitter, and shimmer. However, nonlinear time series analyses associated with complexity and determinism could provide further insights on its regularity and dynamics. This preliminary study investigates the application of phase spaces and recurrence plots to illustrate patterns of vibrato behaviour in solo singing, and how these dynamics can be studied using nonlinear metrics such as sample entropy and recurrence quantification analysis.

Sixty-eight vibrato notes from three operatic pieces sung by Luciano Pavarotti were analysed. Rate, extent, jitter, shimmer, sample entropy, and determinism were calculated for all notes, and phase spaces and recurrence plots created. Results revealed trends and transitions of vibrato behaviour and time-varying characteristics not observable using previous metrics. Classic nonlinear time series analyses methods seem to be promising tools to better understand characteristics of vibrato complexity, which could be valuable to pedagogy and understanding stylistic traits of different genres.

1. INTRODUCTION

Research on vocal vibrato suggests that regularity and periodicity can affect its auditory perception and pleasantness. In one of the earliest empirical studies of music acoustics, Seashore defined vibrato as “a periodic pulsation, generally involving pitch, intensity, and timbre” [1], the word periodic referring to the fact that its waveform can resemble that of a perfectly sinusoidal wave. In terms of vibrato production, parameters are associated with singing aesthetics—a rate of between 5 Hz to 8 Hz and extent of up to 200 cents peak to peak [2]—and research suggests that “samples with better quality are the most periodic ones, . . . more constant and regular over time” [3].

More recently, vibrato regularity has been addressed by jitter and shimmer analyses [4], which are a percentage measure of variability in rate and extent, respectively [5].

However, these features may not thoroughly capture and illustrate the intricacies of the time varying nature of vibrato. Nonlinear time series analyses associated with complexity and periodicity are tools that have been used in voice science, and music rhythm and synchrony [6–15] that could bring further insights on vibrato stability and its production and perception.

Utilising a qualitative approach, phase spaces and recurrence plots are tools that illustrate the dynamics of time series. Phase space reconstruction of time series can be achieved using a time delayed signal, whereby the observation of the geometric pattern is considered the first step in studying the dynamical behaviour [16]. In the phase space, the cycles of vibrato would create orbits bounded between the extent of the undulations, with the density of the orbits affected by its regularity.

Derived from phase spaces, recurrence plots show how the reconstructed trajectory repeats itself. As such, recurrence plots can reveal features of time series and contain structures such as homogeneity, disruptions (white bands), periodic/quasi-periodic patterns, single isolated points, or diagonal lines parallel to the line of identity [17]. These different structures provide information about the dynamics of a signal. For instance, disrupted structures such as *white areas* or *bands* can often be associated with transitions or nonstationarities in a time series.

Recurrence quantification analysis, as the name suggests, quantifies the information provided by recurrence plots using measures like determinism or laminarity. These and other recurrence metrics can collectively (along with other methods such as surrogate techniques) suggest periodicity/regularity of time series [18]. Another metric from the family of nonlinear time series analysis, sample entropy, has also been found to be a robust method to quantify complexity in even short and non-stationary time series [19] and shows “promise as a metric of perceptually relevant voice instabilities” [20].

This preliminary, descriptive study investigates vibrato tones from published musical pieces to question whether (a) phase spaces and recurrence plots are valuable for observing patterns of vibrato behaviour in solo singing, and (b) visual differences or patterns would be reflected in nonlinear features such as sample entropy and recurrence quantification analysis and / or other measures of vibrato regularity.

2. METHODOLOGY

Ethical approval for the research was obtained from the Physical Sciences Ethics Committee of the University of York (reference № Acosta20221114). Sixty-eight notes were selected from three published operatic pieces—La Serenata, 'O Sole Mio, Nessun Dorma—recorded by the world-famous operatic tenor Luciano Pavarotti for vibrato analysis. Voice isolation was performed from wav files of the recordings using iZotope Rx 9, and vibrato notes were manually selected and exported as pitch listings using Pratt software [21]. Figure 1 shows a flow diagram of the process.

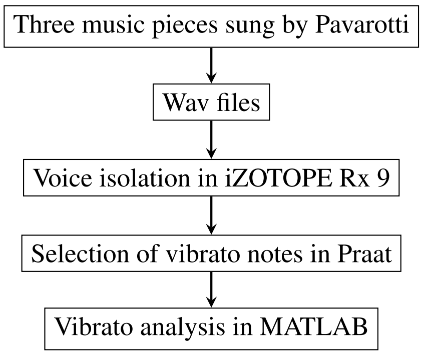


Figure 1. Flowchart of data extraction and analysis

For this study, a note containing vibrato was defined as phonation on a single vowel and single note as defined by the score, with no less than five visible oscillations to allow sufficient length for the analyses. The f_o time series obtained from Pratt had a timestep of 0.01 s ($f_s = 100$ Hz), and each note was imported to MATLAB and smoothed using a low-pass filter with a cutoff frequency of 15 Hz.

A MATLAB script was created to calculate from each note the mean fundamental frequency (Hz), length (s), number of cycles, rate (Hz), extent (cents), jitter, shimmer, sample entropy, and determinism. The values of jitter and shimmer of vibrato were calculated as in Horii 1989 [5]. The SampEn toolbox [22] and CRP toolbox [17] were used to calculate sample entropy and recurrence analysis, respectively.

For the phase space reconstruction of vibrato notes, the embedding dimension was $m = 3$ and embedding delay $\tau = 4$. For sample entropy the tolerance was $r = 0.1$, and for recurrence analysis the adaptative threshold ϵ was set to ensure that the recurrence rate $RR \lesssim 0.05$ [17]. (The RR has an increasing function of ϵ, thus the threshold selected was the maximum possible value—in steps of 0.05—to obtain approximately 0.05 of recurrence rate.)

3. RESULTS

The mean fundamental frequency across notes from the three pieces varied from 175 Hz to 501 Hz, which roughly correspond to the pitches F3 to B4. The numerical analyses of rate, extent, jitter, shimmer, sample entropy, and determinism for all the notes are presented in Table 1. The time series used for the phase space reconstruction (Vibrato[n]) were obtained as the extent of the vibrato, i.e. the deviation from the mean f_o of the studied note in cents. Visual inspection of the phase spaces and recurrence plots revealed observable trends of vibrato behaviour. Figures 2 to 11 present some exemplary notes of phases spaces and recurrence plots that illustrate these trends in dynamics and regularity.

In the phase space plots, the axes from the 3D space correspond to the extent of vibrato and the two delayed signals for the reconstruction, usually contained within -100 and 100 cents. For the recurrence plots, the upper graph shows the f_o time series of the notes, with the horizontal axis corresponding to time (seconds $\times$ 100) and the vertical axis to a normalized version of the extent. (This being that the analysis for recurrence normalizes the signals with a z-score function.) The bottom graphs show the obtained recurrence plot of the signals, which are symmetrically mirrored around the line of identity.

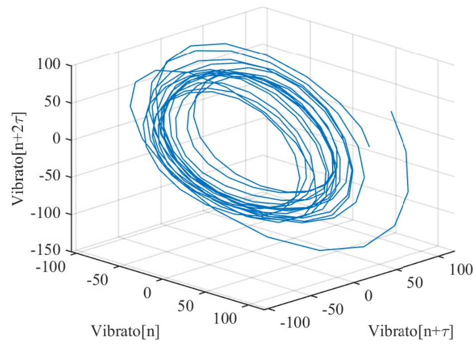


Figure 2. Phase space of note 46.

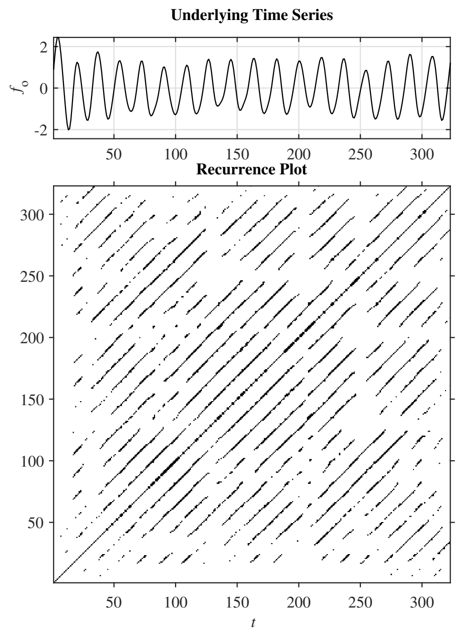


Figure 3. Recurrence plot of note 46.

Figure 2 shows the reconstruction of note 46, which is one of the longest extracted notes, having 19 cycles and a length of 3.3 seconds. The recurrence plot for this note in Figure 3 shows that, with the parameters used in this study, it is possible to detect some of the common structures described by Marwan [17] corresponding to diagonal lines parallel to the line of identity. The same patterns associated with periodicity can be seen in the recurrence plot of note 67, which is also 19 cycles in length (Figure 5), with the main difference that here another structure is present in the form of white bands associated with transitions. This makes it possible to detect three transitory windows in Figure 5 separated by the vibrato cycles that occurred around 140 and 210.

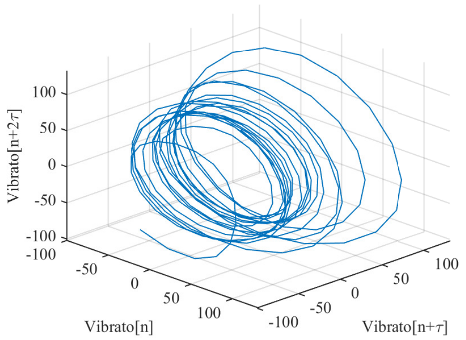


Figure 4. Phase space of note 67.

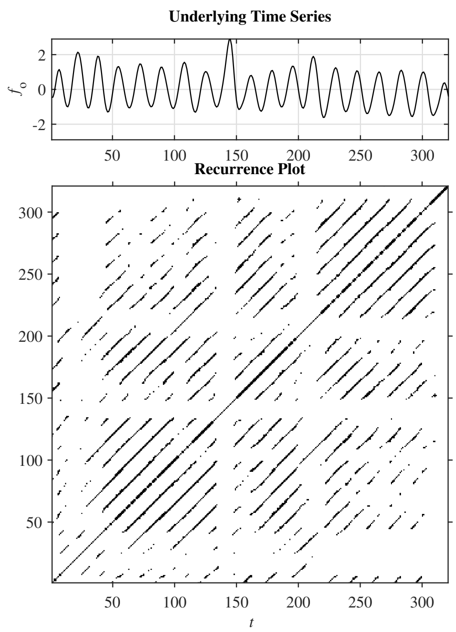


Figure 5. Recurrence plot of note 67.

The results from note 62, with only 5 cycles of vibrato, in contrast to those from the two previous long notes, show that recurrence plots are also suitable for short notes. This can be observed by the point scatters in Figure 7 that reflect the periodicity from the very circular orbit in Figure 6. This phase space reconstruction of the note, along with its low value of sample entropy (SampEn = 0.273) show that these two analyses are also appropriate for short time series.

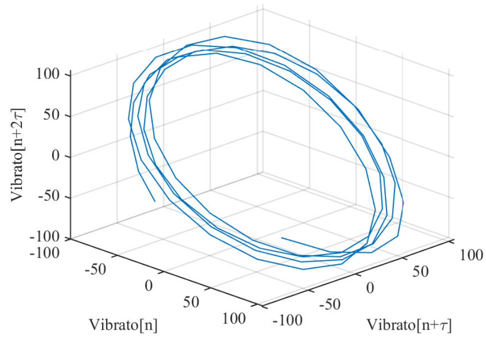


Figure 6. Phase space of note 62.

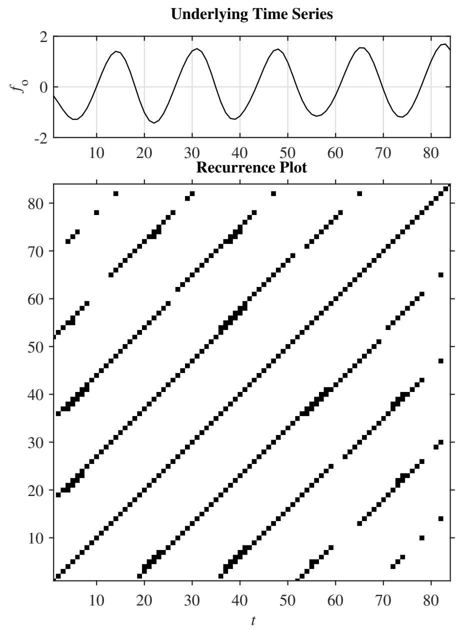


Figure 7. Recurrence plot of note 62.

Figure 8 shows another example of a short vibrato tone. In this case, the note has a high sample entropy value of 1.08 and high determinism of 0.919. As compared to the phase space in Figure 6, the orbits created by this attractor look denser, meaning that the distance between them has higher separation from each other. In other words, the extent varied more over each vibrato cycle. However, it is also observable that the orbits are very smooth, which may be related to its high value of determinism. The recurrence plot of this note (Figure 9) also reflects high complexity, as contrary to note 62, there are fewer diagonal/continuous lines.

Sample entropy values from notes 62 and 17—both approximately one second long—indicate a possible difference in vibrato regularity between the two notes (SampEn = 0.273 for note 62, and SampEn = 1.08 for note 17). However, the measure of determinism did not reflect such difference between the notes ($DET = 0.967$ for note 62, and $DET = 0.919$). This comparison is relevant given that the number of cycles and length of those two notes are very similar, however, the higher sample entropy value of note 17 as well as the visually apparent separation of orbits suggest that note 17 is more irregular than note 62.

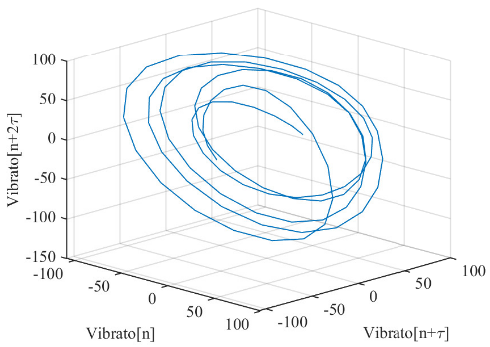


Figure 8. Phase space of note 17.

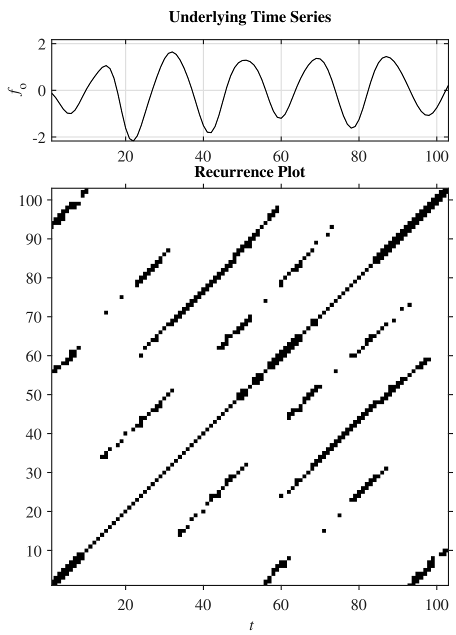


Figure 9. Recurrence plot of note 17.

A number of vibrato tones were excluded from the above analysis because they were not restricted to a single musical note, that is, the singer performed a *portamento* between two adjacent musical notes whilst continuing to phonate on a single vowel and producing vibrato. Examples of such instances are shown in Figure 10 and 11 and illustrate the property of recurrence plots to reveal the windows between transitions or non-stationary events and their potential to allow analysis of transitory behaviour of vibrato across musical notes.

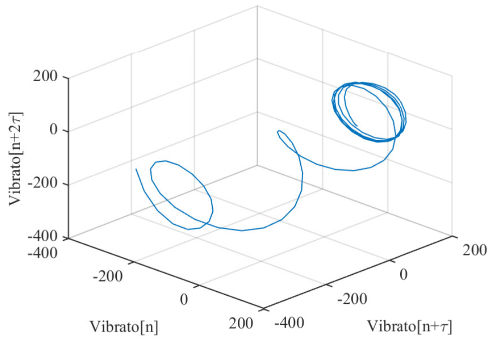


Figure 10. Phase space of note 43.

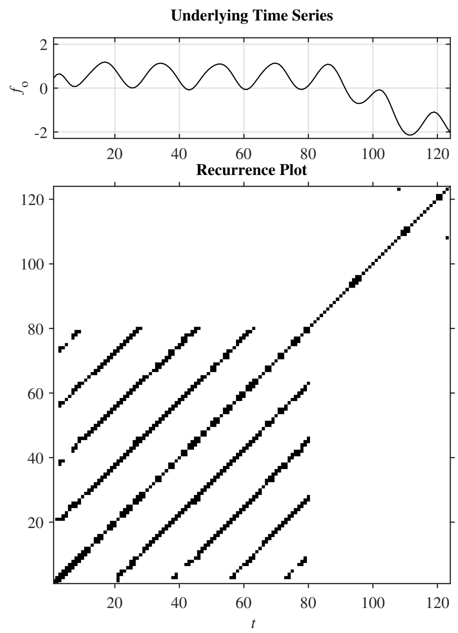


Figure 11. Recurrence plot of note 43.

N	$\overline{f_o}$ (Hz)	Length (s)	Cycles	Rate (Hz)	Extent (cents)	Jitter (%)	Shimmer (%)	SampEn	*DET*
					La Serenata				
1	298	2.83	17	6	117	6.59	22.54	0.558	0.93
2	193	1.6	9	5.6	121	4.63	36.52	0.872	0.908
3	297	2.79	16	5.7	107	6.47	20.97	0.432	0.928
4	297	1.46	9	6.2	121	5.93	35.29	0.542	0.927
5	216	1.72	11	6.4	93	14.49	36.08	0.836	0.869
6	245	2.09	12	5.7	117	3.46	30.16	0.639	0.878
7	394	1.03	6	5.8	128	11.49	27.08	0.371	0.936
8	336	1.03	6	5.8	143	4.57	10.05	0.5	0.859
9	354	1.54	10	6.5	105	7.92	40.69	0.539	0.95
10	332	2.26	13	5.8	91	4.4	32.82	0.701	0.911
11	293	1.28	8	6.3	102	6.98	33.56	0.318	0.912
12	268	1.81	11	6.1	112	4.5	42.21	0.557	0.906
13	195	1.34	7	5.2	149	7.92	22.76	0.768	0.897
14	302	3.26	18	5.5	128	5.12	23.83	0.464	0.92
15	297	2.28	12	5.3	119	7.15	25.97	0.706	0.908
16	300	0.99	6	6.1	117	7.81	23.44	0.478	0.922
17	246	1.1	6	5.5	114	5.49	44.31	1.08	0.919
18	403	1.13	7	6.2	127	3.53	40.69	0.542	0.885
19	371	1.23	7	5.7	125	6.86	21.9	0.816	0.92
20	357	1.44	8	5.6	128	8.27	24.73	0.54	0.953
21	333	1.59	9	5.7	112	9.73	23.45	0.591	0.906
22	291	1.12	7	6.3	89	3.46	46.75	0.467	0.915
					'O Sole Mio				
23	211	0.87	6	6.9	153	4.87	29.32	0.359	0.912
24	175	0.82	6	7.3	113	6.58	45.7	0.737	0.881
25	238	1.32	9	6.8	103	15.06	40.25	0.39	0.915
26	210	1.33	8	6	108	4.7	38.71	0.704	0.921
27	432	1.77	10	5.6	131	2.15	14.74	0.408	0.907
28	321	0.95	6	6.3	128	7.18	31.83	0.448	0.946
29	286	1.6	10	6.2	151	5.92	23.9	0.611	0.896
30	287	0.91	6	6.6	123	7.44	43.38	0.591	0.965
31	318	2.69	14	5.2	129	6.22	16.83	0.392	0.939
32	322	1.63	10	6.1	108	5.96	28.67	0.418	0.964
33	314	0.84	5	6	115	1.93	18.15	0.438	0.877
34	317	0.94	6	6.4	141	7.27	35.12	0.468	0.939
35	319	1.83	12	6.6	119	12.01	21.57	0.491	0.913
36	237	1.24	8	6.5	167	10.42	15.72	0.519	0.928
37	235	0.81	6	7.4	117	15.15	30.38	0.521	0.859
38	235	1.22	8	6.6	103	8.41	28.25	0.964	0.892
39	428	0.84	5	6	136	8.74	58.44	1.362	0.881
40	321	0.91	6	6.6	151	12.05	39.26	0.474	0.899
41	284	0.81	5	6.2	135	11.76	36.38	0.562	0.892
42	316	2.65	15	5.7	144	6.95	20.5	0.449	0.916
43	320	1.31	8	6.1	120	9.05	31.34	0.238	0.94
44	315	1.47	9	6.1	148	15.85	24.44	0.451	0.870
45	474	1.51	9	6	136	4.08	30.83	0.464	0.922
46	432	3.3	19	5.8	130	2.97	15.65	0.415	0.935

Table 1. Descriptive characteristics of the notes analysed, vibrato parameters, and regularity metrics. (Highlighted rows correspond to the notes illustrated in phase spaces and recurrence plots.)

N	$\overline{f_o}$ (Hz)	Length (s)	Cycles	Rate (Hz)	Extent (cents)	Jitter (%)	Shimmer (%)	SampEn	*DET*
					Nessun Dorma				
47	255	0.84	6	7.1	121	10.14	43.73	0.764	0.879
48	253	1.06	7	6.6	97	4.75	52.12	0.674	0.859
49	250	1.33	8	6	106	6.14	20.47	0.411	0.897
50	197	1.29	7	5.4	192	3.6	17	0.487	0.886
51	197	1.16	7	6	149	10.59	34.48	0.479	0.895
52	252	1.3	8	6.2	192	5.93	15.5	0.657	0.913
53	449	1.68	10	6	104	5.05	18.81	0.45	0.897
54	451	0.97	6	6.2	103	1.4	35	0.336	0.948
55	437	0.94	6	6.4	81	1.51	53	0.632	0.913
56	380	2.24	13	5.8	143	3.64	23.98	0.39	0.923
57	219	0.93	6	6.5	158	4.63	38.97	0.383	0.839
58	335	3.21	19	5.9	113	5.43	25.54	0.455	0.927
59	301	0.92	6	6.5	129	5.95	48.82	0.501	0.889
60	299	1.16	7	6	116	5	20.76	0.611	0.912
61	449	0.85	6	7.1	97	8.01	50.81	0.477	0.882
62	381	0.91	5	5.5	179	1.93	4.93	0.273	0.967
63	373	0.73	5	6.8	100	6.25	35.56	0.665	0.895
64	330	1.88	11	5.9	134	4.58	22.55	0.414	0.897
65	304	1.55	9	5.8	158	4.17	10.7	0.568	0.92
66	409	2.27	14	6.2	151	4.99	16.76	0.443	0.945
67	501	3.28	19	5.8	120	5.45	24.64	0.537	0.933
68	455	3.26	19	5.8	118	2.66	24.95	0.426	0.942
Mean±SD				6.1±0.5	125±23				

Table 1. (cont.)

4. DISCUSSION

In this study, we investigated the feasibility to utilise nonlinear time series analyses to better understand vocal vibrato. Sixty-eight notes were obtained from the three music pieces as described in the previous section. The length and number of cycles per note was diverse; minimum five cycles and no maximum, resulting in examples ranging from 5 to 19 cycles. The results for the rate and extent show, as expected, that whilst some variation is present, all the notes analysed fall inside the commonly reported values for rate of 5 to 8 Hz and within a peak to peak extent of 200 cents [23]. The mean and standard deviation (SD) of notes (last row of Table 1) provide information about the variability across notes but no insight of small changes in regularity throughout a note.

Phase spaces, recurrence analysis, and sample entropy provided a different perspective on the treatment of vibrato across notes. Both approaches were sensitive to the selection parameters (embedding dimension and delay, tolerance, and threshold) which were set constant across notes based on standard recommendations [17,22]. As suggested by these results, phase spaces and recurrence plots could be a visual aid in assessing dynamic vibrato behavior and a useful resource in vocal pedagogy. A possible application, for instance, would be the detection of vibrato tones with more than one sinusoid [24].

The presented phase space and recurrence plots showed trends that indicate different types of vibrato dynamics being employed in Pavarotti's singing. Based on the sample entropy and determinism values, Pavarotti produced vibrato notes with different levels of complexity. However, the relationships between jitter and shimmer, entropy, and determinism are not straightforward with no clear trends observed in the current data set. This is because sample entropy / determinism can be sensitive to changes in both rate and extent as well as the shape of the cycles, and they therefore present a different picture of the vibrato behaviour of a note.

A clear advantage of jitter and shimmer over nonlinear metrics can be that these metrics offer information specifically about the rate / extent of vibrato: Jitter reflects only variation of rate, without being affected by variation of extent. On the other hand, shimmer explains only variation of extent. An advantage of the presented integrated regularity metrics is that spotlighting vibrato under nonlinear science can allow for more robust and better established methods for quantification of complexity. And also better understanding of which methods are best suited for short and nonstationary time series from experimental data. For instance, regardless of parameter selection, entropy and recurrence analysis have been favored over other measures like Correlation Dimension or Lyapunov Exponent Estimation [25].

As it has been noted before [18], the value of DET can be affected by smoothing of time series, high values reflecting the smooth shape of the waveform. As such, determinism alone cannot be used to infer nonlinearity in time series. But collectively with jitter, shimmer, and sample entropy, it suggests that vibrato is not perfectly periodic as it has sometimes been referred. And these measures from complex systems may allow the more nuanced characteristic of vibrato to be understood.

One limitation of the study is a lack of statistical analyses within any of the numerical analyses performed. Another is that the patterns and trends presented in this study do not necessarily represent a comprehensive study of all different vibrato scenarios that can occur on different genres, or even different singers. Further studies need to include a more extensive data set of vibrato notes, as well as a closer examination of the relationship between all the metrics studied in this pilot study.

Our understanding of these results, for instance, could benefit from the application of Principal Component Analysis to study which metrics are more relevant in terms of variability and how they interact with each other [26]. A group of the variables that come into play in vibrato production can be added, such as musical intention / expression, prescribed dynamics (intensity), vowel, or musical style. Using dimensionality reduction techniques, groups of notes with similar values on principal components could be clustered or classified to study their perceptual significance. Investigating the perceptual correlates of these nonlinear metrics could lead to applications in pedagogical settings.

These features could also be used to classify individual singers or music styles using larger data sets. Recurrence plots could be used to contrast singers with regular vibrato with other singers with a more irregular pattern—or patients with indications of vocal tremor [27, 28]—possibly striving to show statistical differences of *DET* between the two cases.

Finally, more metrics and surrogate techniques could be used to provide a more robust assessment of nonlinearity and periodicity in the time series. As future work, more comprehensive research can question if the level of vibrato chaos/irregularity could relate to the underlying physiological mechanisms of the voice. As neural mechanisms have been associated to physiological aspects of vibrato [29, 30], it would be of interest to understand what these metrics mean physiologically. For instance, if they potentially point at coupled oscillators (e.g. via different neural feedback loops) within the voice production system.

5. CONCLUSIONS

This study sought to analyse steady-state vibrato tones sung by Luciano Pavarotti in published music material. Classic nonlinear time series analyses methods have been shown here to be promising tools to better understand characteristics of vibrato dynamics, allowing consideration of the time-varying characteristics. Application of these metrics to vocal vibrato could be valuable to pedagogy and understanding stylistic traits of different genres.

Acknowledgments

This research was partially funded by a CONACYT-FINBA scholarship (№ 795214). We thank the anonymous reviewers of this paper for their insightful comments and suggestions.

6. REFERENCES

[1] C. E. Seashore, "The natural history of the vibrato," in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 17, no. 12. National Academy of Sciences, 1931, pp. 623–626.

[2] H. Daffern, "Distinguishing characteristics of vocal techniques in the specialist performance of early music," Ph.D. dissertation, University of York, 2008.

[3] J. A. Diaz and H. B. Rothman, "Acoustical comparison between samples of good and poor vibrato in singers," *Journal of Voice*, vol. 17, no. 2, pp. 179–184, 2003.

[4] A. S. Becker and P. J. Watson, "The use of vibrato in Belt and Legit styles of singing in professional female musical-theater performers," *Journal of Voice*, 2022.

[5] Y. Horii, "Frequency modulation characteristics of sustained /a/ sung in vocal vibrato," *Journal of Speech, Language, and Hearing Research*, vol. 32, no. 4, pp. 829–836, 1989.

[6] H. Kantz and T. Schreiber, *Nonlinear time series analysis*, 2nd ed. Cambridge University Press, 2004.

[7] C. T. Herbst, H. Herzel, J. G. Švec, M. T. Wyman, and W. T. Fitch, "Visualization of system dynamics using phasegrams," *Journal of the Royal Society Interface*, vol. 10, no. 85, p. 20130288, 2013.

[8] W. C. Costa, F. M. Assis, B. G. A. Neto, S. C. Costa, and V. J. D. Vieira, "Pathological voice assessment by recurrence quantification analysis," in *ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living*. IEEE, 2012, pp. 1–6.

[9] A. Selamtzis and S. Ternström, "Analysis of vibratory states in phonation using spectral features of the electroglottographic signal," *The Journal of the Acoustical Society of America*, vol. 136, no. 5, pp. 2773–2783, 2014.

[10] C. Fabris, W. De Colle, and G. Sparacino, "Voice disorders assessed by (cross-) Sample Entropy of electroglottogram and microphone signals," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 920–926, 2013.

[11] M. Little, P. Mcsharry, S. Roberts, D. Costello, and I. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *Nature Precedings*, pp. 1–35, 2007.

[12] J. Neubauer, M. Edgerton, and H. Herzel, "Nonlinear phenomena in contemporary vocal music," *Journal of Voice*, vol. 18, no. 1, pp. 1–12, 2004.

[13] P. Tranchant, E. Scholler, and C. Palmer, "Endogenous rhythms influence musicians' and non-musicians' interpersonal synchrony," *Scientific Reports*, vol. 12, no. 1, p. 12973, 2022.

[14] C. Palmer and A. P. Demos, "Are we in time? How predictive coding and dynamical systems explain musical synchrony," *Current Directions in Psychological Science*, vol. 31, no. 2, pp. 147–153, 2022.

[15] A. Ravignani and P. Norton, "Measuring rhythmic complexity: A primer to quantify and compare temporal structure in speech, movement, and animal vocalizations," *Journal of Language Evolution*, vol. 2, pp. 4–19, 2017.

[16] S. H. Strogatz, *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*, 2nd ed. Westview Press, 2015.

[17] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports*, vol. 438, no. 5-6, pp. 237–329, 2007.

[18] N. Marwan, "How to avoid potential pitfalls in recurrence plot based data analysis," *International Journal of Bifurcation and Chaos*, vol. 21, no. 4, pp. 1003–1017, 2011.

[19] A. Delgado-Bonal and A. Marshak, "Approximate entropy and sample entropy: A comprehensive tutorial," *Entropy*, vol. 21, no. 6, p. 541, 2019.

[20] S. Ternström, D. Johansson, and A. Selamtzis, "FonaDyn—A system for real-time analysis of the electroglottogram, over the voice range," *SoftwareX*, vol. 7, pp. 74–80, 2018.

[21] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9, pp. 341–345, 2001.

[22] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[23] J. Sundberg, "Acoustic and psychoacoustic aspects of vocal vibrato," *STL-QPSR*, vol. 35, pp. 45–68, 1994.

[24] C. T. Herbst, S. Hertegard, D. Zangger-Borch, and P. Å. Lindestad, "Freddie Mercury—acoustic analysis of speaking fundamental frequency, vibrato, and subharmonics," *Logopedics Phoniatrics Vocology*, vol. 42, no. 1, pp. 29–38, 2017.

[25] P. Faure and H. Korn, "Is there chaos in the brain? I. Concepts of nonlinear dynamics and methods of investigation," *Comptes Rendus de l'Académie des Sciences, Sciences de la Vie*, vol. 324, no. 9, pp. 773–793, 2001.

[26] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, 2016.

[27] R. A. Lester-Smith and B. H. Story, "The effects of physiological adjustments on the perceptual and acoustical characteristics of vibrato as a model of vocal tremor," *The Journal of the Acoustical Society of America*, vol. 140, no. 5, pp. 3827–3833, 2016.

[28] D. Abur, R. A. Lester-Smith, A. Daliri, A. A. Lupiani, F. H. Guenther, and C. E. Stepp, "Sensorimotor adaptation of voice fundamental frequency in Parkinson's disease," *PloS One*, vol. 13, no. 1, p. e0191839, 2018.

[29] I. R. Titze, N. P. Solomon, E. S. Luschei, and M. Hirano, "Interference between normal vibrato and artificial stimulation of laryngeal muscles at near-vibrato rates," *Journal of Voice*, vol. 8, no. 3, pp. 215–223, 1994.

[30] P. H. Dejonckere, M. Hirano, and J. Sundberg, *Vibrato*. Singular Publishing Group, 1995.

THE EFFECTS OF MENOPAUSE ON PHONATION AND COLLISION THRESHOLD PRESSURES

Mauro B. Fiuza
Escuela Internacional de Doctorado Universidad Nacional de Educación a Distancia (EIUNED), Madrid, Spain
mauro.barro@edu.uned.es

Filipa M.B. Lã
Facultad de Educación, Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain
filipa.la@edu.uned.es

ABSTRACT

Introduction: Sex steroid hormonal variations associated with menopause have been reported to cause vocal folds' oedema and increased vascularity. These may interfere with vocal fold's motility and their tissue's pliability. To test this hypothesis, phonation and collision threshold pressures (PTP and CTP, respectively) were compared between pre and postmenopausal female professional voice users (FPVUs). Methodology: Audio, electroglottographic and intraoral pressure signals were recorded. Participants were asked to perform diminuendo sequences using the syllable /pa/, repeated at three pitches (A3, E4 and A4). Pressure peaks providing good estimates of subglottal pressure were considered for analysis. The lowest peaks causing vocal fold vibration and vocal fold contact were averaged for each pitch, providing measurements of PTP and CTP, respectively. Results: No significant differences between pre and postmenopausal singers and teachers of singing were found for both PTP and CTP and the three analysed pitches. Discussion and Conclusions: Postmenopausal hypoestrogenism seems not to affect vocal folds' mobility in FPVUs.

1. INTRODUCTION

At menopause, i.e., the end of a female's reproductive life, concentrations of steroid hormones (i.e., oestrogens, progesterone, and testosterone) are significantly reduced, especially for oestrogens. This hypoestrogenism has been related to voice changes; a decrease in fundamental frequency of about 1 to 2 ST has been observed in post as compared to premenopausal females [1]. Possible explanations have included alterations in vocal folds' mucosa, such as oedema and micro varices [2], conditions both interfering with vibratory patterns and the pliability of the vocal folds' mucosal tissues. For example, elevated concentrations of oestrogens and progesterone associated with pregnancy had an effect in rising both PTP and CTP values [3].

The present study aims at investigating whether changes in concentrations of sex steroid hormones associated with menopause, namely hypoestrogenism, may interfere with vocal folds' motility, measured by means of phonation and collision threshold pressures (PTP and CTP, respectively). To test this hypothesis, PTP and CTP were measured as the minimum air pressure required to initiate vocal folds' vibration and contact, respectively, and then compared between pre and postmenopausal female professional voice users (FPVUs). Higher values would reflect less pliable vocal folds' and thus a greater vocal effort [4]

2. METHOD

2.1 Participants

Ethical approval was obtained from Ethical Committees of the Universidad Nacional de Educación a Distancia (UNED) and of the Hospital Clínico San Carlos (CEIC) (ref. 20/674-E), both in Madrid, Spain. Sixty-four healthy female professional voice users (FPVUs) were recruited, namely teachers of different disciplines, including singing, and also professional singers (classical and commercial contemporary music). These professional groups were specifically targeted because for them, even mild voice alterations may hamper the quality of their work [5]. Recruitment was made through: (i) authors' personal cont[3]acts; (ii) e-mails to professional choirs and music conservatoires; (iii) UNED voice Lab social media profiles; (iv) authors' personal social media; (v) Facebook groups of singers and casting in Madrid, Spain; (vi) direct messages to teachers of singing on online music teachers' platforms. FPVUs were allocated into pre and postmenopausal groups according to both clinical and endocrinological criteria, following the recommendations reached by the menopausal staging system workshop [6]: (i) premenopausal group - presence of regular menstrual cycles or, if irregular, not surpassing more than 3 consecutive months of amenorrhea; (ii) postmenopausal group - ≥ 12 months of amenorrhea, without surprising 5 consecutive years, and concentrations of follicle stimulating hormone (FSH) more than 30 IU/ L and concentrations of oestrogens less than 25 ρg/ mL. All participants were non-smoking healthy volunteers. They all provided information on professional voice use experience, vocal health and voice use habits, and hormonal and reproductive status. In addition, blood samples were collected for determination of concentrations of gonadotropins (FSH and luteinising hormone, LH) and sex steroid hormones.

From a total of 64 FPVUs initially recruited, 13 were excluded according to the following exclusive criteria: (i) not being a schoolteacher, a teacher of singing or a professional singer (n = 4); and (ii) being at perimenopausal stage (n = 7). In addition, two more participants needed to be excluded from analysis due to poor-quality electroglottographic (EGG) signals. Furthermore, data from three other participants were also excluded from analysis as reliable intraoral pressure peaks could not be used as good estimates of subglottal pressure (P_{sub}). This yielded a total of

48 FPVUs for data analysis, 23 pre (44 yrs. ± 2.6 SD), and 25 postmenopausal (54.5 yrs ± 3.5 SD).

2.2 Tasks

Participants were asked to perform diminuendo sequences with the syllable /pa/ while having a thin tube inserted into the corner of their mouth. This allowed the measurement of intraoral pressure during the occlusion of the consonant /p/ as an estimate of P_{sub} [7]. They started at comfortable loudness, progressively reducing it until unvoiced /pa/ syllables were reached. Participants were instructed to perform the sequences in legato to facilitate the production of flat intraoral pressure peaks, required for reliable estimates of P_{sub}. The task was performed in three pitches, A3 (± 220 Hz), E4 (± 333 Hz) and A4 (± 440 Hz) and repeated at least five times. Performing diminuendo sequences maintaining the pitch is a difficult task, especially for untrained participants. Thus, the task was visually displayed as a voice map using the costume made free software FonaDyn (Sten Ternström, Sweden) whenever necessary. The voice map showed frequency [Hz] and sound pressure level (SPL) [dB] in real-time, so that participants could target the required pitch, maintaining it while varying sound level throughout the whole task, as discussed elsewhere [8].

2.3 Equipment and procedures

Recordings were made at a room the Hospital Clínico San Carlos in Madrid, Spain, inside a sound treated acoustic environment. Audio, electroglottographic and intraoral pressure signals were simultaneously recorded using the software FonaDyn, version 2.4 (by Sten Ternström, Sweden)[9]. Signals were synchronised using an external soundcard Fireface UCX (RME Audio, Germany) connected to two AC/DC interfaces, ES-6 and ES-3 (Expert Sleepers Limited, UK) via two optical cables, and to a PC, via a USB cable. Audio signals were captured using two microphones connected to the Fireface UCX sound card. The EGG signal was collected using an EG2-PCX 2 device (Glottal Enterprises, USA), while the intraoral pressure peaks via a PS-100 Subglottal Pressure Monitor (Glottal Enterprises, USA). Both devices were connected to the ES-6 interface and their signals were visually inspected via an ES-3 interface connected to an oscilloscope (Data Mordax Systems, USA) (see Figure 1).

Audio signals were recorded using two microphones: (i) an omnidirectional condenser measurement ECM8000 microphone (Behringer, Germany), placed at 30 cm from the front of the mouth of each participant; and (ii) an omnidirectional 4066 condenser headset microphone (DPA Microphones, Denmark), placed at 5 cm from the centre of the mouth.

Audio signals were calibrated using a sound level calibrator (Extech Instruments, Germany), following the recommendations described elsewhere [10]. The procedure was as follows: the measurement microphone was inserted tightly into the sound level calibrator to avoid sound leaks while the calibrator generated a stable sinusoidal 1 kHz tone with an SPL of 94 dB. The gain of the measurement microphone was adjusted to that same SPL. Then the calibrator was removed. The participant was asked to phonate a sustained /a/ vowel at a constant loudness and pitch while adjusting the gain of the headset microphone so that it matched the SPL of the measurement microphone.

The EGG signal was captured via two electrodes held externally on the thyroid notch by means of an adjustable neck collar. Contact gel was priorly applied to facility electric conductivity.

Intraoral pressure peaks were measured by inserting one end of a small plastic tube into the corner of the participant's mouth during the /p/ occlusion. The other end of the plastic tube was connected to the PS-100 monitor. Pressure peaks were calibrated in cmH_2O by previously recording the pressure peak caused by the water displacement when inserting the tube into a bottle containing a known hight of water.

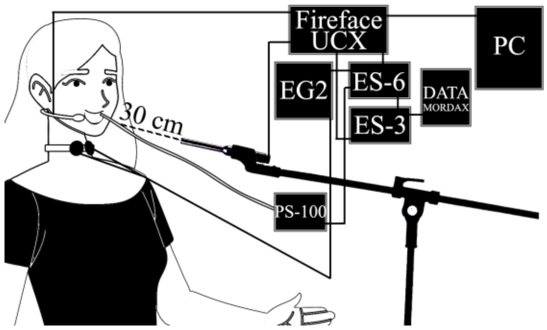


Figure 1. Recording setup.

2.4 Analysis

Audio, EGG, and intraoral pressure peaks were analysed using the costume made software Sopran (by Svante Granqvist, Sweden). Figure 2 displays the three signals used to guide and measure PTP and CTP: from top to bottom, audio, derivative of the EGG signal (dEGG) after being squared, and intraoral pressure peaks. The squaring of dEGG was made using the *Process* module in the costume made free software Sopran (Svante Granqvist, Sweden). The squared dEGG offers a clearer visual inspection of presence or absence of vocal folds' contact as compared to dEGG, and thus allows a better identification of intraoral pressure peaks causing and not causing vocal fold's contact [4]. CTP was extracted as the average between the last peak causing contact (1) and the first that failed to cause contact (2). PTP was calculated as the average between the last pressure peak producing voicing (3), as observed from the audio channel, and the first peak failing to produce voice (4) [4].

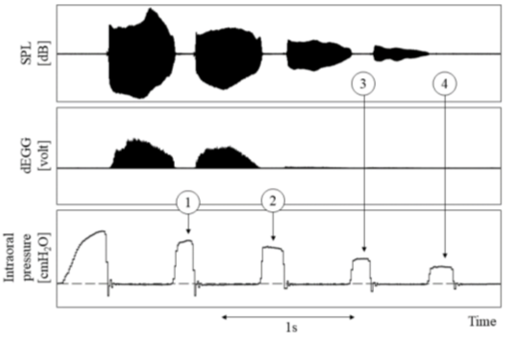


Figure 2. Audio, derivative EGG signal (dEGG) squared, and intraoral pressure signals. Numbers 1 and 2 indicate pressure peaks used to measure PTP, and 3 and 4 those used to measure CTP.

A Kolmogorov-Smirnov test was carried out to determine data distribution. Comparisons between pre and postmenopausal FPVUs were made using either a Chi-Squared test of a student's t-test, for normally distributed categorical or continuous variables, respectively, or a Mann-Whitney U test for those variables showing a skewed distribution.

3. RESULTS

3.1 Participants' characteristics

Pre and postmenopausal FPVUs groups in terms of age by means a t-Student test and in terms of voice education and body mass index (BMI) using a Chi-square test. A Mann-Whitney U test was carried out to compare groups as to what concerns professional experience and professional voice use. Results are summarized in Table 1.

	Pre-menopause (n = 23)	Post-menopause (n = 25)	*p*
Age	44 ± 2.6*	54.5 ± 3.5*	< .000§
Days from last menstrual cycle	17 (10-23) †	1096 (731-1513)†	< .000§
Profession			.807
Teachers	13 (56.5) ‡	15 (60) ‡	
Singers	10 (43.5) ‡	10 (40) ‡	
Professional experience [yrs.]	15 (14-18) †	30 (22-32) †	<.000§
Professional voice use [hours/ day]	4 (4-5) †	4 (4-5) †	.528
Voice education			.076
No	10 (43.5) ‡	9 (36) ‡	
Yes, in the past	0	5 (20) ‡	
Yes, currently	13 (56.5) ‡	11 (44) ‡	
BMI			.479
Underweight	1 (4.3) ‡	2 (8) ‡	
Healthy	14 (60.9) ‡	11 (44) ‡	
Overweight	5 (21.7) ‡	10 (40) ‡	
Obese	3 (13) ‡	2 (8) ‡	

Table 1. Participants' characteristics. * Mean ± Standard deviation. † Median (interquartile range). ‡ n (%), §statistically significance.

As expected, postmenopausal FPVUs showed significantly higher age, number of days from last menstrual cycle and years of professional experience as compared to premenopausal FPVUs. In addition, postmenopausal FPVUs also showed a tendency for a higher prevalence of voice education.

3.2 Phonation Threshold Pressure and Collision Threshold Pressure

Figure 3 displays PTP as boxplots comparing pre and postmenopausal FPVUs. PTP was marginally greater for pre as compared to postmenopausal FPVUs for all pitches, except for E4 (postmenopausal FPVUs, A3 = 4.08, E4 = 5.02, and A4 = 7 cmH_2O; premenopausal group, A3 = 3.74, E4 = 5.24, and A4 = 6 cmH_2O). However, significant differences were not reached between groups, as showed by the results of a Mann-Whitney U Test ($p > 0.05$).

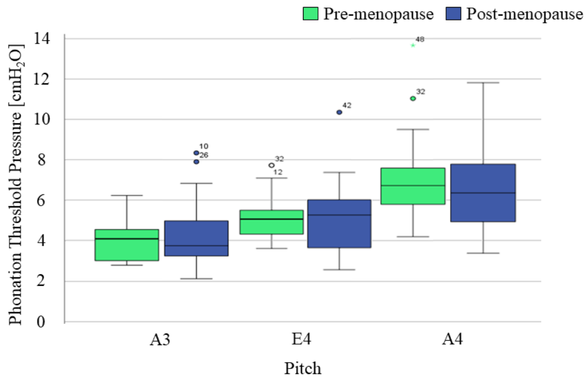


Figure 3. Boxplots comparing Phonation Threshold Pressures between pre and postmenopausal FPVUs at pitches A3, E4 and A4.

As to what concerns CTP, also significant differences could not be found between groups ($p > 0.05$). However, as displayed in boxplots of Figure 4, postmenopausal FPVUs showed a trend for higher CTP as compared to premenopausal FPVUs for all three pitches. For the postmenopausal group, CTP median values were 5.43, 7.27 and 9 cmH_2O, for A3, E4 and A4, respectively; for the premenopausal group, A3, E4 and A4 showed CTP of 4.99, 6.69 and 8 cmH_2O, respectively.

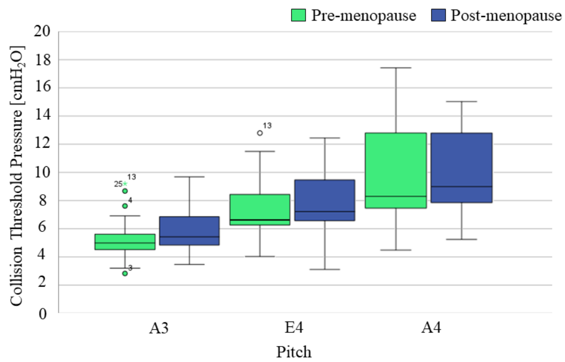


Figure 4. Boxplots comparing Collision Threshold Pressures between pre and postmenopausal FPVUs at pitches A3, E4 and A4.

4. DISCUSSION

The depletion of concentrations of sex steroid hormones during menopause, especially oestrogens, have been associated with vocal folds swelling and oedema [11]. In this study, PTP and CTP were measured to evaluate whether the significant decrease in concentrations of oestrogens (and elevated concentrations of FSH) could affect vocal folds mobility in FPVUs.

For A3, PTP was 4.08 and 3.74 cmH_2O for pre and postmenopausal FPVUs, respectively. These values are within the ± 4 cmH_2O found for healthy female non-singers' voices around 200 Hz (close to A3) [12]. Such results corroborate that our participants had no signs of alterations in vocal folds' mobility. However, a large variability in data distribution was observed in Figures 3 and 4. Thus, despite the absence of statistical significance, individual differences can be observed and should not be disregarded [13]. One should not ignore the complains of some women, who report voice difficulties during menopause [5],

For PTP, maximum values (excluding outliers) are always higher for post as compared to premenopausal FPVUs. PTP for participants with the highest levels reached approximately 3, 2 and 6 cmH_2O more than the median values found for A3, E4 and A4, respectively. In these cases, such difference could account for greater vocal effort. As with respect to CTP, maximum values were higher for postmenopausal participants and for pitches A3 and E4. For A4, the maximum was considerably smaller, but the median was greater. Such great variability corroborates previous work suggesting that, for a small number of individuals, menopause may cause great vocal distress.

A possible limitation of this study could be the inclusion of both trained and non-trained voices. Singers and teachers of singing may have learnt to circumvent effects of mild changes in vocal folds' pliability due to mild oedema and micro varices associated with menopausal changes, thus concealing effects of menopause on both PTP and CTP. However, this investigation was the first assessing the impacts of menopause on PTP and CTP and therefore a larger sample size (n = 48) seemed to be worthwhile to include. In addition, age difference between pre and postmenopausal could have limited our results, as age can also interfere with the pliability of vocal folds' mucosal tissues and therefore PTP and CTP. To circumvent this possible limitation, a longitudinal study following perimenopausal FPVUs PTP and CTP throughout years of menopausal transition would be ideal, although difficult to pursue in terms of participants compliance to longitudinal evaluations that can take up to five years as menopausal transitional periods may last.

The current investigation was the first study that has included both clinical and endocrinological criteria to inform on accurate distribution of participants on pre and postmenopausal groups. Future research can complement these results by looking at other metrics related to vibratory characteristics of the vocal folds in a wider range of frequencies and intensities, comparing, for example, voice maps of pre and postmenopausal FPVUs.

5. CONCLUSIONS

The hypoestrogenism associated with menopause seems to not interfere with vocal folds' mobility for pitches A3, E4 and A4 of FPVUs. Further investigations are needed to identify the extent to which menopause may cause changes in vocal folds' vibratory patterns in a wider range of frequencies.

Acknowledgments

The authors would like to thank all participants, Dr. Pluvio Coronado and the staff of the Gynaecology and Obstetrics Department of Hospital Clínico San Carlos. This work was supported by the Programa de Atracción de Talento Investigador de la Comunidad de Madrid [Ref. 2018T1/HUM-12172] and Becas Santander-UNED 2022.

6. REFERENCES

[1] F. M. B. Lã and D. Ardura, "What Voice-Related Metrics Change With Menopause? A Systematic Review and Meta-Analysis Study," *Journal of Voice*, 2020, doi: 10.1016/j.jvoice.2020.06.012.

[2] J. Abitbol, P. Abitbol, and B. Abitbol, "Sex hormones and the female voice," *Journal of Voice*, vol. 13, no. 3, pp. 424–446, 1999, doi: 10.1016/S0892-1997(99)80048-4.

[3] F. M. B. Lã and J. Sundberg, "Pregnancy and the singing voice: Reports from a case study," *Journal of Voice*, vol. 26, no. 4, pp. 431–439, 2012, doi: 10.1016/j.jvoice.2010.10.010.

[4] L. Enflo, J. Sundberg, C. Romedahl, and A. Mcallister, "Effects on vocal fold collision and phonation threshold pressure of resonance tube phonation with tube end in water," *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 5, pp. 1530–1538, 2013, doi: 10.1044/1092-4388(2013/12-0040).

[5] M. B. Fiuza, M. L. Sevillano, and M. B. L. Filipa, "Singing at Menopause: A Systematic Review with Pedagogical Implications," *Musicae Scientiae*, p. 102986492110478, 2021, doi: 10.1177/10298649211047884.

[6] S. D. Harlow *et al.*, "Executive summary of the stages of reproductive aging workshop + 10: Addressing the unfinished agenda of staging reproductive aging," *Journal of Clinical Endocrinology and Metabolism*, vol. 97, no. 4, pp. 1159–1168, 2012, doi: 10.1210/jc.2011-3362.

[7] S. Hertegård, J. Gauffin, and P. Å. Lindestad, "A comparison of subglottal and intraoral pressure measurements during phonation," *Journal of Voice*, vol. 9, no. 2, pp. 149–155, 1995, doi: 10.1016/S0892-1997(05)80248-6.

[8] F. M. B. Lã and M. B. Fiuza, "Real-Time Visual Feedback in Singing Pedagogy: Current Trends and Future Directions," *Applied Sciences (Switzerland)*, vol. 12, no. 21, 2022, doi: 10.3390/app122110781.

[9] S. Ternström, D. Johansson, and A. Selamtzis, "FonaDyn — A system for real-time analysis of

the electroglottogram, over the voice range," *SoftwareX*, vol. 7, pp. 74–80, 2018, doi: 10.1016/j.softx.2018.03.002.

[10] J. G. Švec and S. Granqvist, "Guidelines for selecting microphones for human voice production research," *Am J Speech Lang Pathol*, vol. 19, no. 4, pp. 356–368, 2010, doi: 10.1044/1058-0360(2010/09-0091).

[11] E. D'Haeseleer, H. Depypere, S. Claeys, F. L. Wuyts, S. de Ley, and K. M. van Lierde, "The impact of menopause on vocal quality," *Menopause*, vol. 18, no. 3, pp. 267–272, 2011, doi: 10.1097/gme.0b013e3181f3ee36.

[12] I. R. Titze, "Phonation threshold pressure: A missing link in glottal aerodynamics," *Journal of the Acoustical Society of America*, vol. 91, no. 5, pp. 2926–2935, 1992, doi: 10.1121/1.402928.

[13] A. Sand, "Inferential Statistics Is an Unfit Tool for Interpreting Data," *Applied Sciences*, vol. 12, no. 15, p. 7691, 2022, doi: 10.3390/app12157691.