IMPLEMENTATION OF A CONTINUOUSLY VARIABLE DELAY LINE BY CROSSFADING BETWEEN SEVERAL TAP DELAYS

Thibaut CARPENTIER (thibaut.carpentier@ircam.fr) [1]

[1]**STMS Lab, IRCAM — CNRS — Sorbonne Université — Ministère de la Culture**, 1, place Igor-Stravinsky, Paris, 75004 France

ABSTRACT

Delay lines are ubiquitous in computer music applications: they are used to create audio effects or to simulate the sound propagation of moving sources. When one wants to synthesize a delay which varies over time, two main strategies are typically implemented: a) by using a "fractional" delay line, which simulates a non-integer delay based on an interpolator filter; or b) by applying a crossfade between the initial delay and the new desired delay value. These two techniques may induce pitch shift and/or spectral artifacts which are not tolerable in certain applications such as spatial reproduction by wavefield synthesis (WFS). In this article we propose a new method for creating a continuously variable delay line; the technique is an extension of the crossfade delay technique which exploits a superposition of several "auxiliary" tap delays whose times and gains are determined by a methodology similar to fractional delay interpolating filters. We show that it is thus possible to reduce coloration artifacts, at the expense of a higher computational cost.

1. INTRODUCTION

Delay lines (DL) are ubiquitous in signal processing and computer music applications. They are notably used for the production of audio effects such as vibrato, flanging, chorus, phasing [1] [2, Chapter 2] [3, Chapter 2] [4, Chapter 7.8], for physical modeling synthesis of musical instruments [5–7], for sound spatialization [8, 9] and artificial reverberation [10, 11], and for the simulation of propagation of moving sources, as observed in the Doppler effect or Leslie cabinet [12–14] or in auralization systems [15, Section 3.2] [16, Fig. 22] [17], etc. They also play a central role for sample rate conversion [18, 19], beamforming [20], and synchronization in wireless communication systems [21].

In many use cases, it is necessary to vary the delay value over time. When delay lines are implemented in the digital domain, the delay variation requires special attention. A "basic" delay line is generally implemented as a circular buffer, with two pointers which represent the read and write positions respectively, and which are incremented at each time step. In the digital domain, the read and write positions are necessarily integer values (expressed in number of samples). It is known that, for such a "basic" delay line, the transition from a delay τ_1 to a delay τ_2 (where τ_1 and τ_2 are rounded to the nearest integer values) induces a discontinuity in the output signal, which manifests itself as audible artifacts such as clicks or "zipper noise" [3, Chapter 2] [7].

To overcome this problem, it is necessary to implement an interpolation process in order to smooth the variations and minimize (or eliminate) the discontinuities. Two main strategies are listed in the literature, and very widely used in practice: a) implement a fractional delay line, using an interpolator filter to simulate non-integer delay values; b) perform a crossfade between two read pointers corresponding to the delays τ_1 and τ_2. These approaches each have limitations — which will be discussed in more detail in the following sections: a) a fractional delay line induces a pitch shift comparable to a Doppler effect; depending on the use cases, this transposition can be artistically relevant or undesirable; b) the crossfade technique generates a comb filtering effect with detuning (undesired distorsion of the magnitude and phase spectrum) which can be more or less pronounced depending on the situation.

Spatial reproduction of sound sources by wavefield synthesis (WFS) [9, 22] is a use case of interest, and the initial motivation of this study. In WFS, each loudspeaker (called "secondary source") of the reproduction system contributes to recreating the wave front emanating from a virtual source (or "primary source"), and the driving functions of the loudspeakers — derived from the Kirchhoff-Helmholtz and Rayleigh I integral equations — involve delay lines (a priori with fractional delay values). Each time the virtual source moves, the relative delays of all the speakers should vary continuously. Given that the number of secondary sources is generally large (several dozens to hundreds) and that the range of the delays can be considerable (up to several tens of milliseconds depending on the configuration and source movements), WFS reproduction is particularly sensitive to the delay line artifacts discussed previously. The pitch changes of a fractional delay and the spectral colorations of a crossfade delay are both undesirable [19, 23–26].

In this article we propose a new method for creating a continuously variable delay line; the technique is an extension of the crossfade delay that exploits "auxiliary" tap delays whose times and gains are determined by a methodology similar to fractional delay interpolating filters. We will show that it is therefore possible to reduce coloration artifacts, at the expense of a higher computational cost.

The article is organized as follows: first, we present elementary reminders on fractional delay lines (Section 2), and crossfading delay lines (Section 3). In Section 4, we present and evaluate the proposed method, before concluding.

2. FRACTIONAL DELAY LINE (FDL)

2.1 Ideal response

Fractional delay lines (FDL) allow to simulate any delay $\tau \in \mathbb{R}^+$. The delay value is decomposed into an integer part and a fractional part: $\tau = \tau_i + \beta$ where $\tau_i = \lfloor \tau \rfloor$ is the integer part and $\beta \in [0, 1) = \tau - \tau_i$ is the fractional part. For an input sequence $x[n]$ originating from the discretization of a band-limited analog signal $x(t)$, the output of a fractional delay line is written $y[n] = x[n - (\tau_i + \beta)]$. This value a priori lies "between two samples", which is not possible. Instead, one must determine the discrete values (on the sampling grid) by interpolation.
Many methods have been proposed and extensively discussed (see in particular [1, 5, 27–29] [4, Chapter 5.5]). These techniques generally rely on the approximation of an "ideal delay operator" by a finite impulse response filter (FIR) or infinite impulse response filter (IIR). Considering the relation $Y(z) = z^{-\tau} X(z)$, the ideal transfer function is written in the frequency domain:

$$H_{\text{ideal}}\left(e^{i\omega}\right) = e^{-i\omega\tau} , \tag{1}$$

where $z = e^{i\omega}$, and $\omega = 2\pi f / f_s$ denotes the angular frequency. As a result, the impulse response of the system writes $\forall n$

$$h_{\text{ideal}}[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_{\text{ideal}}\left(e^{i\omega}\right) e^{i\omega n}\, d\omega = \text{sinc}\,(n - \tau) \tag{2}$$

where $\text{sinc}\,(x) = \frac{\sin(\pi x)}{\pi x}$ is the cardinal sine function.
When the desired delay τ is integer, the response reduces to a single pulse at the discrete time $n = \tau$, as in "basic" DL. When τ is real and non-integer, the impulse response $h_{\text{ideal}}[n]$ has an infinite length, and is non-causal. It is therefore not suitable for real-time applications. The FDL must be implemented with an approximate solution.

2.2 Approximations of the ideal solution

A large number of approximation techniques have been proposed in the literature. The simplest and most intuitive approach relies on approaching the ideal delay operator by linear interpolation between adjacent samples $x[n-1]$ and $x[n]$. This technique has a low computational cost, and it provides qualitatively satisfactory results when the spectrum of the signal x is mainly concentrated in low frequencies. To process large-band signals, it is preferable to apply oversampling, however this considerably increases the computational cost.

Another simple way to approximate the ideal operator h_{ideal} is to truncate the reponse, and shift it temporally to ensure causality. This results in an FIR filter of order M. It is known that the impulse response of this filter is subject to the Gibbs phenomenon, and consequently its magnitude spectrum exhibits ripples (see for example [4, Figure 5.46]), which is generally undesirable.

Another widely used approach consists of implementing the FDL using a 1st order all-pass IIR filter [27] [4, Chapter 5.5.3], known as 1st order Thiran filter [30] [29, Section 20.4.2.2.], which guarantees a maximally flat group delay for $\omega \to 0$. The approach has several advantages: the solution is obtained analytically, the computation cost is very low, and the frequency response is flat for all frequencies. However, the phase delay approximation is only valid in a domain restricted to low frequencies, and degrades rapidly at higher frequencies. Furthermore, the impulse response is relatively long, which is problematic for transient sounds, when the input signal or the all-pass coefficient varies quickly (causing audible clicks) [5, 31]. Alternative implementations have been proposed to mitigate this problem [25,32,33], but at the expense of higher complexity.

Finally, another commonly used strategy approximates the desired function by an interpolator polynomial of order N [27] [4, Chapter 5.5.4]. The simplest analytical solution is provided by the Lagrange polynomial. The coefficients of the FIR Lagrange interpolator can be found e.g. in [27] [29, Section 20.3.4.]. For $N = 1$, it is equivalent to linear interpolation; for $N \to \infty$, it converges to the ideal sinc operator [7]. The filter exhibits a low-pass behavior, more pronounced for even orders N. As a consequence, order $N = 3$ is frequently used, as it offers a tradeoff between moderate computational cost, and acceptable magnitude/phase response — the frequency response is almost flat up to $f_s/4$, see Figure 2(a). Figure 1 displays an example impulse response of the Lagrange interpolator filter of order $N = 3$.

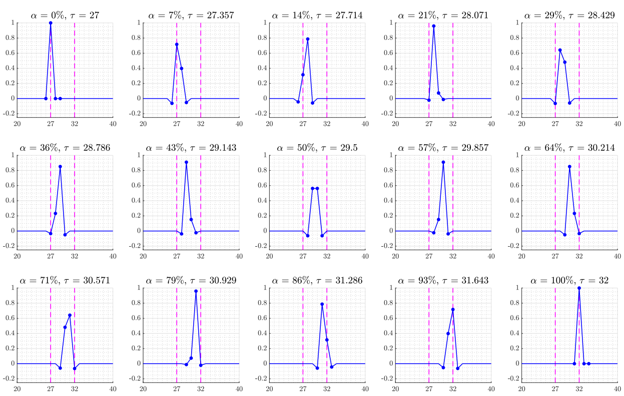

Figure 1. Impulse response of a FDL for $\tau_1 = 27$, $\tau_2 = 32$, and for different values of the interpolation factor α between 0% and 100%. The x-axis and y-axis represent time (in samples) and amplitude respectively. The purple dashed lines symbolize τ_1 and τ_2. The FDL is implemented with 4-point Lagrange interpolation ($N = 3$).

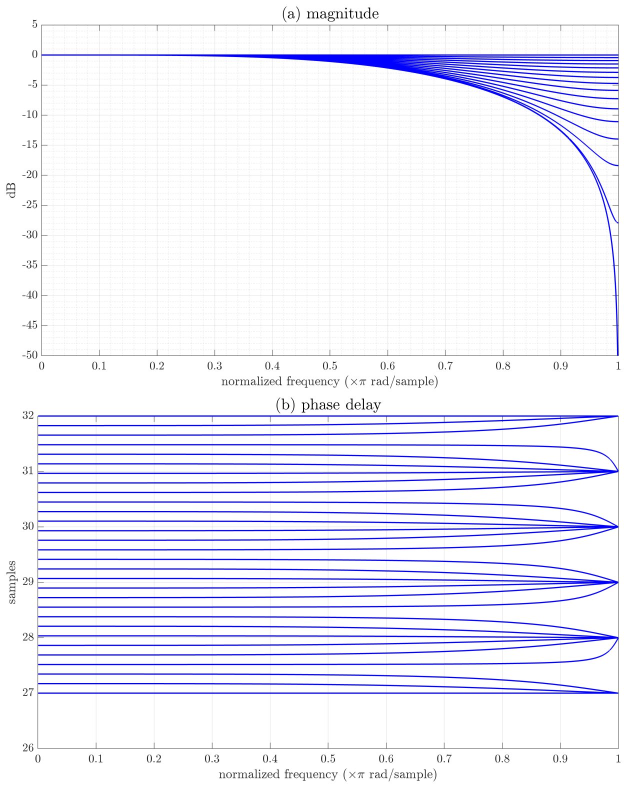

Figure 2. (a) Frequency response and (b) phase delay of a FDL with Lagrange interpolation. The simulated parameters are similar to Figure 1. The frequency response is most attenuated when $\alpha = 50\%$, with $\left\| H\left(e^{i\omega}\right)\big|_{\omega=\pi} \right\| = 0$.

2.3 Artifacts when varying the delay value

The accuracy of the sinc ideal response approximation is generally improved by using higher order interpolation schemes, however the computational cost then becomes prohibitive for real-time applications. Furthermore, whatever the method used to implement the FDL, audible artifacts are generated when the delay value varies continuously over time such that $y(t) = x\big(t - \tau(t)\big)$. For the sake of simplicity, we assume for example that $\tau(t)$ varies linearly between two integer delays τ_1 and τ_2. Without loss of generality, we will also assume that $\tau_2 > \tau_1$. We note α the "interpolation" factor such that, at each time t, the (fractional) delay τ is worth

$$\tau(t) = \alpha\,\tau_1 + (1-\alpha)\,\tau_2 \;, \tag{3}$$

with $\alpha \in \mathbb{R}$ and $0 \leq \alpha \leq 1$. Let $\Delta = \tau_2 - \tau_1$ be the delay variation, and δ be the interpolation duration to go from τ_1 to τ_2, i.e. $\alpha\big|_{t=0} = 1$ and $\alpha\big|_{t=\delta} = 0$.

By varying τ, a momentary transposition is generated; this is analogous to the Doppler effect and results in a clearly audible pitch change [1, 7] [34, Chapter 7.7] [35]. This Doppler effect is physically correct: the transition from τ_1 to τ_2 amounts to changing the physical distance which separates a listener from a sound source (the delay line realistically "modeling" the acoustic propagation channel). When $\Delta > 0$, the transposition is downward; and vice versa when $\Delta < 0$ [34, 36]. As noted by Puckette [34, Chapters 7.7 and 7.8], this effect is quite pronounced even when the interpolation duration (δ) is significantly greater than the delay variation (Δ); the effect of pitch shift is for example clearly audible even with long duration $\delta > 100\,|\Delta|$ (By comparison, when we consider amplitude variations, with gain ramps, satisfactory results can generally be obtained even with relatively short interpolation times, e.g. $\delta \approx 20$ ms for $|\Delta| \approx 100$ dB).

In some applications, the induced Doppler effect may be acceptable or even desirable. In fact, this justifies the ubiquitous use of delay lines in audio effects implementing pitch variations or modulations (vibrato effect, for example) [34, Chapter 7.9] [29, Section 20.7.4.] [1, 37].
On the other hand, in spatialization or auralization contexts, the Doppler shift – even though it simulates a realistic auditory effect – is sometimes undesirable, considered disruptive and harmful to timbral content [19, 23, 38]. As mentioned in the introduction, this is all the more critical for WFS rendering, if the virtual sources move "rapidly" (generating variations $\Delta \gg 1$, of the order of a few tens of milliseconds or more), and if the number of loudspeakers is high.
Similarly, in auralization systems, delay lines are used to simulate the acoustic channel (i.e. the reflection paths) between a sound source and a listener; when one or the other moves, the delays must be continually adjusted. Depending on the simulated geometric configuration, the number of simulated delays (acoustic reflection paths) can be significant; the constraints are therefore similar to the WFS rendering presented above.

3. CROSSFADE BETWEEN 2 TAP DELAYS (XDL)

3.1 Principle

When the Doppler pitch shift is undesirable, or when $\Delta \gg 1$, the alternative method of crossfading delay line (XDL) may be used. This approach consists of using a delay line with two tap delays (i.e. two read pointers), corresponding to τ_1 and τ_2 respectively. To simulate a variation of $\tau(t)$ over time, a crossfade is applied i.e. the amplitude of the two tap delays is interpolated for a duration δ. For the sake of simplicity, a linear crossfade is commonly used, but other types of transition curves can also be implemented.
The XDL technique is rarely discussed in the literature, probably because it is conceptually very simple, and not relying on any physical foundation. Nevertheless, it is briefly mentioned in several publications and patents, for instance concerning the creation of legato effect in the waveguide synthesis of string instruments [39–42]. It is also mentioned in [7, Paragraph "Large Delay Changes"] [43].
Despite its low representation in the academic literature, the XDL approach is very commonly used in computer music tools. For example, it is implemented under the name "Double Delay with Interpolation (DDI)" or "sdelay" (*smooth delay*) in the `Faust` libraries [44]. The crossfade of tap delays is also used (under the name "xdelay") in the WFS implementation of `Spat` [45] and in the mixing

workstation `Panoramix` [46]. In the Max environment, several implementations can be found such as `vdb~` (by Benjamin Thigpen), `ej.vdb~` (by Emmanuel Jourdan), or `M4L.vdelay~` (included in `Max for Live`).

3.2 Implementation

The software implementation of the XDL method is straightforward, and does not raise any particular challenges. However, several remarks can be made — see also [7]. a) A single delay line is sufficient, with one write pointer and two read pointers; thus the memory space requirement is comparable to an FDL. b) During the crossfade period, the computation cost is doubled. c) The duration δ of the crossfade should be long enough to achieve smooth results. However, there is no general rule for choosing δ as a function of Δ, the quality of the rendering may depend on the nature of the input signal. d) The XDL approach is compatible with non-integer τ_1 and τ_2 delays; in this case, the read pointers will use one of the fractional interpolation techniques presented in Section 2. To guarantee an artifact-free result, it may then be wise to take into account the length of the transient response of the interpolator filter: for example, when using an order N filter, one can “pre-warm” the filter for $N+1$ samples before starting the crossfade operation. However, if the duration of the fade δ is large compared to the duration of the transient response of the interpolator filter, the “pre-warming” step is not essential (the transient response of the filter is not audible).

3.3 Limitations

Figure 3 shows the impulse response of an XDL, for different values of the interpolation factor α between 0% and 100%. While the system performs a kind of morphing between the two pure delays τ_1 and τ_2, it does not achieved any interpolation, i.e. it does not synthesize any actual intermediate delay values.

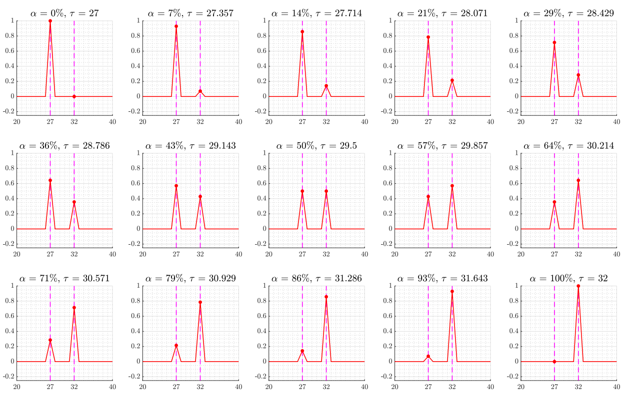

Figure 3. Impulse response of an XDL, for $\tau_1 = 27$, $\tau_2 = 32$, and for various values of the interpolation factor α between 0% and 100%. The x-axis and y-axis represent time (in samples) and amplitude respectively. The vertical purple dashed lines symbolize τ_1 and τ_2.

The transfer function of the system is

$$H(z) = \alpha\, z^{-\tau_1} + (1-\alpha)\; z^{-\tau_2} = \alpha\, z^{-\tau_1} \left(\frac{z^{\Delta} + \frac{1-\alpha}{\alpha}}{z^{\Delta}} \right) . \tag{4}$$

In the righthand term we recognize the typical response of a feedforward comb filter [4, Chapter 5.2]. The frequency response is shown in Figure 4(a); it exhibits notches for the pulsations $\omega = \frac{\pi}{\Delta}, \frac{3\pi}{\Delta}, \frac{5\pi}{\Delta}, \ldots$ The (linear) amplitude of these notches is zero when $\alpha = 50\%$. The phase delay (Figure 4(b)) also exhibits some distortions.
In fact, XDL crossfading is subject to the typical comb filtering coloration. The perception of this spectral coloration may be affected by several factors (input signal, Δ, δ). Depending on the target applications, the artifacts can be tolerated (or sometimes inaudible), or they may be deemed unacceptable.

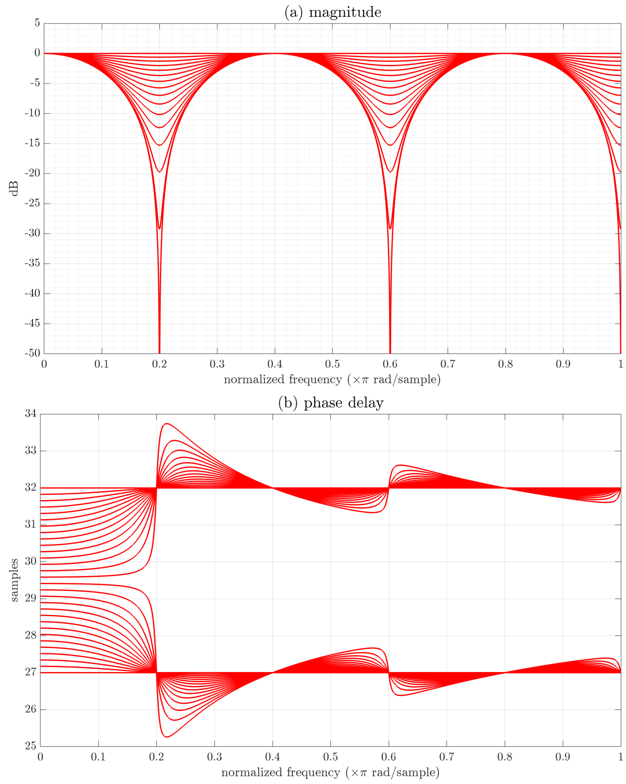

Figure 4. (a) Frequency response and (b) phase delay of an XDL for $\tau_1 = 27$, $\tau_2 = 32$, and for different values of the interpolation factor α between 0% and 100%. The frequency response shows periodic notches for angular frequencies $\frac{\omega}{\pi} = \frac{1}{\Delta} = 0.2$, $\frac{\omega}{\pi} = \frac{3}{\Delta} = 0.6$, and $\frac{\omega}{\pi} = \frac{5}{\Delta} = 1$.

4. PROPOSED METHOD: CROSSFADE BETWEEN $(2K+2)$ TAP DELAYS

4.1 Principle

In order to overcome the limitations of the two previous approaches, we propose a new method which is somehow a compromise between FDL and XDL. We take the XDL as a starting point; instead of reading and fading between two tap delays τ_1 and τ_2, we propose to add other auxiliary tap delays in order to "temper" the excessively strong comb filtering effect. It is then a matter of choosing the discrete times τ_k and the amplitudes of these auxiliary delays over time. For simplicity, we assume that the start and end delays τ_1 and τ_2 are integers. It seems obvious that the auxiliary delays must be symmetric with respect to τ_1 and τ_2, as there is no reason to favor one or the other. To limit the computational cost, it is desirable that the auxiliary delays t_k be integer values (but this is not a necessary condition). Finally, for the sake of simplicity, it is preferable that the delays t_k are regularly sampled (it is not proven that this criterion is necessary). Therefore, we suggest using a "sampling step" Δ, so that the delay values be $\{\tau_1 - k\Delta\}$ and $\{\tau_2 + k\Delta\}$, with $k = 0, 1, ..., K$. In other words, the delays t_k write

$$\begin{cases} t_k = \tau_1 - (K-k)\,\Delta\,, \text{ if } 0 \leq k \leq K \\ t_k = \tau_2 + (k-K-1)\,\Delta\,, \text{ if } (K+1) \leq k < (2K+2)\,. \end{cases} \tag{5}$$

This results in $(2K+2)$ tap delays in total. Note that the XDL presented in Section 3 is a particular case corresponding to $K = 0$.

We then need to determine the crossfading law, that is to say, for each time t, the amplitude associated with each delay t_k. By analogy with the reasoning of Section 2.1, we ideally seek a solution in the form of a sinc function, reaching its maximum at $t = \tau$, however with a "width" Δ:

$$h_{\text{ideal}}(t) = \text{sinc}\left(\frac{t-\tau}{\Delta}\right)\,. \tag{6}$$

In the discrete-time domain, we can therefore approximate the ideal response by truncation:

$$0 \leq k < (2K+2),\ \ h[k] = \text{sinc}\left(\frac{t_k-\tau}{\Delta}\right)\,. \tag{7}$$

This principle is illustrated in Figure 5, for $K = 4$.

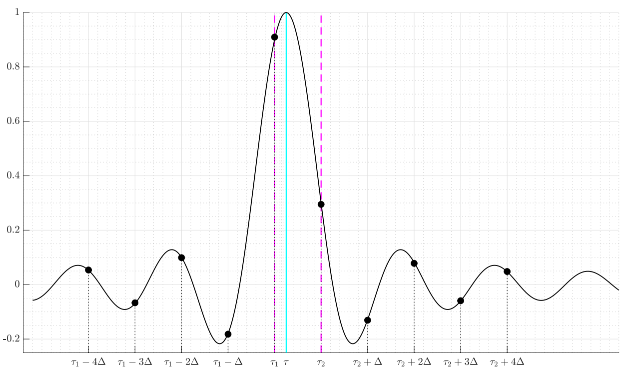

Figure 5. Principle of the proposed method: approximation of the ideal impulse response by truncation. Example with 10 tap delays (i.e. $K = 4$). The vertical purple dashed lines symbolize τ_1 and τ_2. The cyan line symbolizes $\tau(t)$, here for $\alpha = 25\%$. The solid black line depicts $\text{sinc}\left(\frac{t-\tau}{\Delta}\right)$. The black markers represent $h[k]$ for $0 \leq k < (2K+2)$.

For causality reasons, it is required that $\tau_1 - k\Delta \geq 0$, which puts a constraint on K such that $K \leq \left\lfloor \frac{\tau_1}{\Delta} \right\rfloor$.

Furthermore, if we note L the allocated length of the delay line, the constraint $\tau_2 + k\Delta \leq L$ must be verified, hence imposing: $K \leq \left\lfloor \frac{(L-\tau_2)}{\Delta} \right\rfloor$. This leads to:

$$K \leq \min\left(\left\lfloor \frac{\tau_1}{\Delta} \right\rfloor, \left\lfloor \frac{(L-\tau_2)}{\Delta} \right\rfloor\right)\,. \tag{8}$$

4.2 Results

The impulse response of the proposed system is displayed in Figure 6 for different values of the interpolation factor α ranging from 0% to 100%. The behavior is analogous to Figure 3, however with auxiliary "pulses", some of which exhibit a negative amplitude. Note that the amplitude of the two main delays (for $t = \tau_1$ and $t = \tau_2$) is different from the Figure 3.

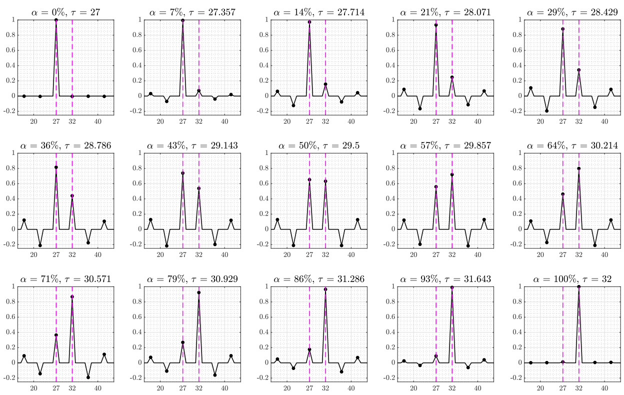

Figure 6. Impulse response of the proposed method (here with $K = 2$), for $\tau_1 = 27$, $\tau_2 = 32$, and for different values of the interpolation factor α between 0% and 100%. The x-axis and y-axis represent time (in samples) and amplitude respectively. The purple dashed lines symbolize τ_1 and τ_2.

Figure 7 illustrates the frequency response of the system, for different values of the interpolation factor α between 0% and 100%. The phase delay generally behaves like the XDL (Figure 4(b)). The magnitude spectrum exhibits notches for the same angular frequencies $\omega = \frac{\pi}{\Delta}, \frac{3\pi}{\Delta}, \frac{5\pi}{\Delta}, ...,$ but these notches are significantly narrower than for XDL (Figure 4(a)). In other words, the undesirable comb filtering effect is reduced. Apart from these notches, we also observe slight ripples around 0 dB, which are undesired; however the range of these ripples remains moderate (less than 2 dB in this example).

4.3 Influence of the number of tap delays

It is obvious that the number K plays a preponderant role. By increasing K (within the authorized range Eq. 8), the spectral effect of comb filtering is reduced. To visualize this effect, we superimpose in Figure 9 the frequency curves for different values of K, and in the worst-case scenario $\alpha = 50\%$. We observe that an increase of K significantly reduces the width of the notches.

Figure 8 presents similar results, now in the case $\alpha = 25\%$;

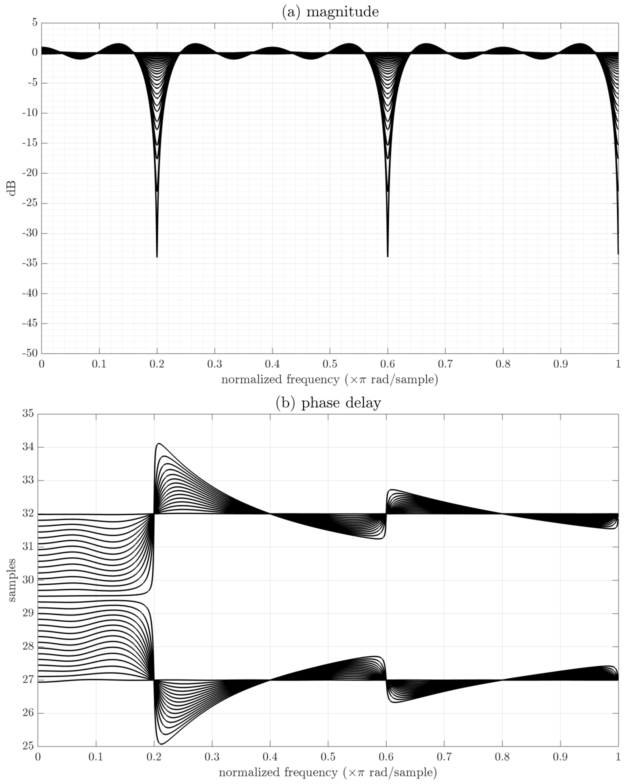

Figure 7. (a) Frequency response and (b) phase delay for the proposed method for $K = 2$, $\tau_1 = 27$, $\tau_2 = 32$, and for different values of the interpolation factor α between 0% and 100%.

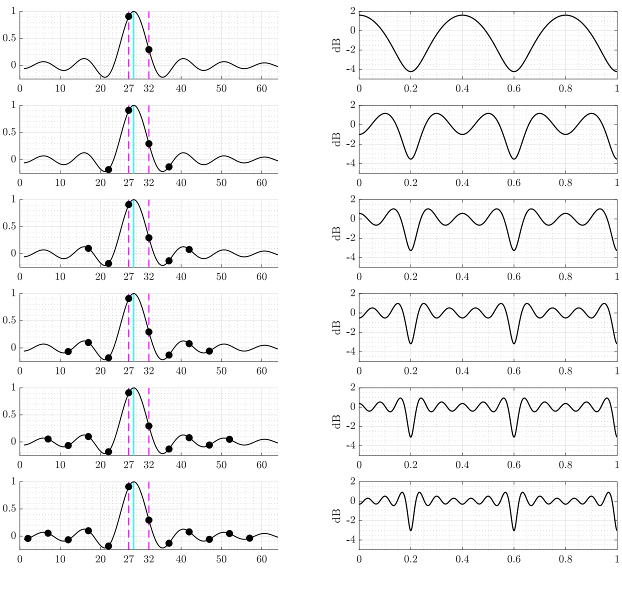

Figure 8. The curves are for $\tau_1 = 27$, $\tau_2 = 32$, and $\alpha = 25\%$ (i.e. $\tau = 28.25$ samples). Left column: impulse response for the proposed method. In solid black line, the function $\mathrm{sinc}\left(\frac{t-\tau}{\Delta}\right)$. The cyan vertical line symbolizes $\tau = 28.25$. Right column: frequency response. First line: $K = 0$; Second line: $K = 1$; Third line: $K = 2$; etc.

it also shows the impulse response for $K = 0, 1, ..., 5$.

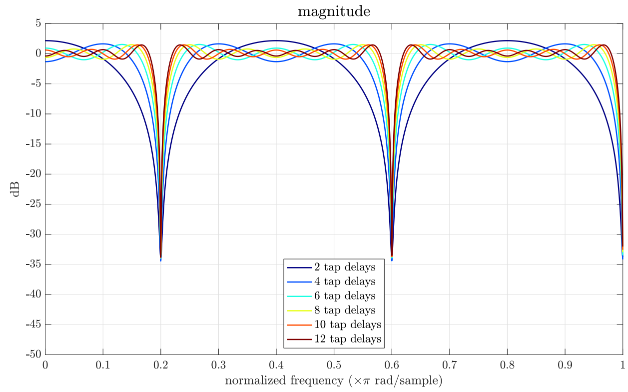

Figure 9. Frequency response of the proposed method, for $\tau_1 = 27$, $\tau_2 = 32$, $\alpha = 50\%$, and $K = 0, 1, 2, ..., 5$.

The computational cost of the proposed method is obviously proportional to the total number of delays $(2K + 2)$. The choice of K is necessarily a compromise. For WFS applications, where the number of delay lines may be high (depending on the number of primary and secondary sources), it is preferable to restrict to $K = 1$, which already offers an audible improvement compared to XDL (based on informal listening).

5. CONCLUSION

In this article we proposed a new method for creating a continuously variable delay line. When the delay varies over time, we operate a crossfade between several “auxiliary” read pointers. The times and gains of these auxiliary tap delays are chosen so as to minimize the typical comb filtering coloration issue of XDL. The introduction of auxiliary tap delays, however, has a significant impact on the computational cost. In future work, a more precise evaluation of the method’s performance remains to be carried out, in particular in comparison with alternative approaches such as “samplerate conversion” [18, 19, 24, 47, 48], or Lagrange interpolator filter implemented in the form of a Farrow structure [21, 49].
Likewise, it would be necessary to conduct a perceptual study of residual artifacts.

6. REFERENCES

[1] J. Dattorro, “Effect Design : Part 2: Delay-Line Modulation and Chorus,” *Journal of the Audio Engineering Society*, vol. 45, no. 10, pp. 764 – 788, October 1997. https://www.aes.org/e-lib/browse.cfm?elib=10159

[2] U. Zölzer, *DAFX: Digital Audio Effects*, 2nd ed. John Wiley & Sons, March 2011. https://doi.org/10.1002/9781119991298

[3] J. D. Reiss and A. P. McPherson, *Audio Effects: Theory, Implementation and Application*, 1st ed. Boca Raton, FL, USA: CRC Press, January 2015. https://doi.org/10.1201/b17593

[4] A. Uncini, *Digital Audio Processing Fundamentals*, 1st ed. Cham, Switzerland: Springer International Publishing, February 2022. https://doi.org/10.1007/978-3-031-14228-4

[5] V. Välimäki, "Discrete-Time Modeling of Acoustic Tubes Using Fractional Delay Filters," Ph.D. dissertation, Helsinki University of Technology, Helsinki, Finland, December 1995.

[6] S. Bilbao, *Numerical Sound Synthesis : Finite Difference Schemes and Simulation in Musical Acoustics*. John Wiley & Sons, October 2009. https://doi.org/10.1002/9780470749012

[7] J. O. Smith, *Physical Audio Signal Processing: for Virtual Musical Instruments and Digital Audio Effects*. W3K Publishing, December 2010. https://ccrma.stanford.edu/~jos/pasp/

[8] F. Rumsey, *Spatial Audio*, 1st ed. New York, NY, USA: Routledge, June 2001. https://doi.org/10.4324/9780080498195

[9] S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, and F. Zotter, "Spatial Sound With Loudspeakers and Its Perception: A Review of the Current State," *Proc. of the IEEE*, vol. 101, no. 9, pp. 1920 – 1938, September 2013. https://doi.org/10.1109/jproc.2013.2264784

[10] W. G. Gardner, "Reverberation Algorithms," in *Applications Of Digital Signal Processing To Audio And Acoustics*, M. Kahrs and K. Brandenburg, Eds. New York, NY, USA: Springer, 2002, ch. 3, pp. 85 – 131. https://doi.org/10.1007/0-306-47042-X_3

[11] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty Years of Artificial Reverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 5, pp. 1421 – 1448, July 2012. https://doi.org/10.1109/tasl.2012.2189567

[12] J. O. Smith, S. Serafin, J. S. Abel, and D. Berners, "Doppler Simulation and the Leslie," in *Proc. of the 5th International Conference on Digital Audio Effects (DAFx)*, Hamburg, Germany, September 2002.

[13] R. Kronland-Martinet and T. Voinier, "Real-Time Perceptual Simulation of Moving Sources: Application to the Leslie Cabinet and 3D Sound Immersion," *EURASIP Journal on Audio, Speech, and Music Processing*, pp. 1 – 10, 2008. https://doi.org/10.1155/2008/849696

[14] Z.-H. Fu and T. Zhu, "Moving Sound Image Rendering in Realistic Audio Effects," in *Intelligent Science and Intelligent Data Engineering*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 722 – 730. https://doi.org/10.1007/978-3-642-31919-8_92

[15] T. Lokki, "Physically-based Auralization - Design, Implementation, and Evaluation," Ph.D. dissertation, Helsinki University of Technology, Espoo, Finland, November 2002.

[16] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating Interactive Virtual Acoustic Environments," *Journal of the Audio Engineering Society*, vol. 47, no. 9, pp. 675 – 705, September 1999. http://www.aes.org/e-lib/browse.cfm?elib=12095

[17] P. Schäfer, J. Fatela, and M. Vorländer, "Interpolation of scheduled simulation results for real-time auralization of moving sources," *Acta Acustica*, vol. 8, p. 9, 2024. https://doi.org/10.1051/aacus/2023070

[18] A. Tarczynski, W. Kozinski, and G. Cain, "Sampling rate conversion using fractional-sample delay," in *Proc. of IEEE ICASSP*, Adelaide, SA, Australia, April 1994. https://doi.org/10.1109/icassp.1994.390042

[19] A. Franck, "Efficient Algorithms for Arbitrary Sample Rate Conversion with Application to Wave Field Synthesis," Ph.D. dissertation, Universitätsverlag Ilmenau, Ilmenau, Germany, 2012.

[20] P. Murphy, A. Krukowski, and A. Tarczynski, "An efficient fractional sample delayer for digital beam steering," in *Proc. of the IEEE ICASSP*, Munich, Germany, April 1997, pp. 2245 – 2248. https://doi.org/10.1109/icassp.1997.599498

[21] C. Farrow, "A continuously variable digital delay element," in *Proc. of the IEEE International Symposium on Circuits and Systems*, Espoo, Finland, June 1988. https://doi.org/10.1109/iscas.1988.15483

[22] R. Rabenstein, S. Spors, and P. Steffen, "Wave Field Synthesis Techniques for Spatial Sound Reproduction," in *Topics in Acoustic Echo and Noise Control*, E. Hänsler and G. Schmidt, Eds. Springer, 2006, ch. 13, pp. 517 – 545. https://doi.org/10.1007/3-540-33213-8_13

[23] A. Franck, A. Graefe, T. Korn, and M. Strauss, "Reproduction of Moving Sound Sources by Wave Field Synthesis: An Analysis of Artifacts," in *Proc. of the AES 32nd International Conference: DSP For Loudspeakers*, Hillerød, Denmark, September 2007. https://www.aes.org/e-lib/browse.cfm?elib=14206

[24] A. Franck, K. Brandenburg, and U. Richter, "Efficient Delay Interpolation for Wave Field Synthesis," in *Proc. of the 125th Convention of the Audio Engineering Society (AES)*, San Francisco, CA, USA, October 2008. https://www.aes.org/e-lib/browse.cfm?elib=14764

[25] F. Winter and S. Spors, "On fractional delay interpolation for local wave field synthesis," in *Proc. of the 24th European Signal Processing Conference (EUSIPCO)*, Budapest, Hungary, August 2016. https://doi.org/10.1109/eusipco.2016.7760682

[26] J. Ahrens and S. Spors, "Reproduction of Moving Virtual Sound Sources with Special Attention to the Doppler Effect," in *Proc. of the 124th Convention of the Audio Engineering Society (AES)*, Amsterdam, Netherlands, May 2008. http://www.aes.org/e-lib/browse.cfm?elib=14493

[27] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine, "Splitting the Unit Delay," *IEEE Signal Processing Magazine*, vol. 13, no. 1, pp. 30 – 60, January 1996. https://doi.org/10.1109/79.482137

[28] V. Välimäki and T. I. Laakso, "Principles of fractional delay filters," in *Proc. of the IEEE ICASSP*, Istanbul, Turkey, June 2000. https://doi.org/10.1109/ICASSP.2000.860248

[29] ——, "Fractional Delay Filters – Design and Applications," in *Nonuniform Sampling*, 1st ed., F. Marvasti, Ed. New York, NY, USA: Springer New York, November 2001, ch. 20, pp. 835 – 895. https://doi.org/10.1007/978-1-4615-1229-5_20

[30] J.-P. Thiran, "Recursive digital filters with maximally flat group delay," *IEEE Transactions on Circuit Theory*, vol. 18, no. 6, pp. 659 – 664, 1971. https://doi.org/10.1109/tct.1971.1083363

[31] S. A. Van Duyne, D. A. Jaffe, G. P. Scandalis, and T. S. Stilson, "A Lossless, Click-free, Pitchbend-able Delay Line Loop Interpolation Scheme," in *Proc. of the International Computer Music Conference (ICMC)*, Thessaloniki, Greece, September 1997. http://hdl.handle.net/2027/spo.bbp2372.1997.068

[32] V. Välimäki, T. I. Laakso, and J. MacKenzie, "Elimination of Transients in Time-Varying Allpass Fractional Delay Filters with Applications to Digital Waveguide Modeling," in *Proc. of the International Computer Music Conference (ICMC)*, Banff, Canada, September 1995. http://hdl.handle.net/2027/spo.bbp2372.1995.096

[33] V. Valimaki and T. Laakso, "Suppression of transients in variable recursive digital filters with a novel and efficient cancellation method," *IEEE Transactions on Signal Processing*, vol. 46, no. 12, pp. 3408 – 3414, 1998. https://doi.org/10.1109/78.735314

[34] M. S. Puckette, *The Theory and Technique of Electronic Music*. World Scientific Press, May 2007. https://doi.org/10.1142/6277

[35] D. Griesinger, "Improving Room Acoustics Through Time-Variant Synthetic Reverberation," in *Proc. of the 90th Convention of the Audio Engineering Society (AES)*, Paris, France, February 1991. http://www.aes.org/e-lib/browse.cfm?elib=5679

[36] J. Pakarinen, M. Karjalainen, V. Valimaki, and S. Bilbao, "Energy behavior in time-varying fractional delay filters for physical modeling synthesis of musical instruments," in *Proc. of the IEEE ICASSP*, vol. 3, Philadelphia, PA, USA, March 2005. https://doi.org/10.1109/ICASSP.2005.1415631

[37] V. Lazzarini, J. Timoney, and T. Lysaght, "Adaptive FM Synthesis," in *Proc. of the 10th International Conference on Digital Audio Effects (DAFx)*, Bordeaux, France, September 2007.

[38] J. Jaworski and W. L. Martens, "A Novel Method for Simulating Movement in Multichannel Reverberation Providing Enhanced Listener Envelopment," in *Proc. of the AES International Conference on Spatial Reproduction - Aesthetics and Science*, Tokyo, Japan, August 2018. http://www.aes.org/e-lib/browse.cfm?elib=19660

[39] D. A. Jaffe and J. O. Smith, "Performance Expression in Commuted Waveguide Synthesis of Bowed Strings," in *Proc. of the International Computer Music Conference (ICMC)*, Banff, Canada, September 1995, pp. 343 – 346. http://hdl.handle.net/2027/spo.bbp2372.1995.098

[40] ——, "Patent WO1998010355A1 – Digital signal processing system and method for generating musical legato using multitap delay line with crossfader," Palo Alto, CA, USA, September 1997.

[41] J. O. Smith, "Physical Modeling Synthesis Update," *Computer Music Journal*, vol. 20, no. 2, p. 44, 1996. https://doi.org/10.2307/3681331

[42] S. A. V. Duyne, D. A. Jaffe, G. P. Scandalis, , and T. S. Stilson, "Patent 5 742 532 – System and method for generating fractional length delay lines in a digital signal processing system," April 1998.

[43] K. Bogdanowicz and R. Belcher, "Using Multiple Processors for Real-Time Audio Effects," in *Proc. of the AES 7th International Conference: Audio in Digital Times*, Toronto, ON, Canada, May 1989, pp. 337 – 342. http://www.aes.org/e-lib/browse.cfm?elib=5449

[44] J. O. Smith, "Signal Processing Libraries for Faust," in *Proc. of the Linux Audio Conference (LAC)*, Stanford, CA, USA, April 2012.

[45] T. Carpentier, "Spat: a comprehensive toolbox for sound spatialization in Max," *Ideas Sónicas*, vol. 13, no. 24, pp. 12 – 23, June 2021. https://hal.science/hal-03356292

[46] ——, "A versatile workstation for the diffusion, mixing, and post-production of spatial audio," in *Linux Audio Conference*, Saint-Etienne, France, May 2017. https://hal.archives-ouvertes.fr/hal-01527754

[47] A. Franck and K. Brandenburg, "An overall optimization method for arbitrary sample rate converters based on integer rate SRC and lagrange interpolation," in *Proc. of the IEEE WASPAA*, New Paltz, NY, USA, October 2009. https://doi.org/10.1109/aspaa.2009.5346472

[48] D. Rocchesso, "Fractionally addressed delay lines," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 6, pp. 717 – 727, 2000. https://doi.org/10.1109/89.876310

[49] A. Franck, "Efficient Algorithms and Structures for Fractional Delay Filtering Based on Lagrange Interpolation," *Journal of the Audio Engineering Society*, vol. 56, no. 12, pp. 1036 – 1056, December 2008. https://www.aes.org/e-lib/browse.cfm?elib=14647