# A social network integrated game experiment to relate tapping to speed perception and explore rhythm reproduction

**Guillaume Bellec**
ENSTA Paristech
bellec@ensta.fr

**Anders Elowsson, Anders Friberg**
Royal Institute of Technology,
Speech, Music and Hearing
{afriberg,elov}@kth.se

**Daniel Wolff, Tillman Weyde**
City University London,
Department of Computer Science
{daniel.wolff.1,t.e.weyde}@city.ac.uk

## ABSTRACT

During recent years, games with a purpose (GWAPs) have become increasingly popular for studying human behaviour [1–4]. However, no standardised method for web-based game experiments has been proposed so far. We present here our approach comprising an extended version of the CaSimIR social game framework [5] for data collection, mini-games for tempo and rhythm tapping, and an initial analysis of the data collected so far. The game presented here is part of the Spot The Odd Song Out game, which is freely available for use on Facebook and on the Web [1].

We present the GWAP method in some detail and a preliminary analysis of data collected. We relate the tapping data to perceptual ratings obtained in previous work. The results suggest that the tapped tempo data collected in a GWAP can be used to predict perceived speed. I toned down the above statement as I understand from the results section that our data are not as good as When averaging the rhythmic performances of a group of 10 players in the second experiment, the tapping frequency shows a pattern that corresponds to the time signature of the music played. Our experience shows that more effort in design and during runtime is required than in a traditional experiment. Our experiment is still running and available on line.

## 1. INTRODUCTION

Collecting perceptual data from listening experiments is a tedious task and the resulting data sets are typically small (tens or hundreds of entries). On the other hand, in music information retrieval (MIR), the size of music collections has exceeded 10 million songs (20m in the Spotify Library [2], 12m in iTunes store in 2010 [3]). To also gather perceptual data on music on a larger scale, the concept of games with a purpose (GWAPs), as defined in 2006 by von Ahn [6], has been applied in some recent MIR projects [1–3,5]. In comparison to traditional experiments, the number of participants in a GWAP can be very large at low cost (TagATune reached 14442 unique players [1]). However, the design and evaluation of GWAPs requires more effort than traditional experiments, as there is less control over the experimental conditions and no human interaction with the subject during the experiment.

Social networks have reached world wide popularity in a relatively short time. Facebook was founded in 2004, and had one billion monthly active users in December 2012. Integrating a GWAP into social networks is thus an opportunity to reach potentially large numbers of players and to gather contextual information.

This paper presents the tempo and rhythm sections of the *Spot The Odd Song Out* Facebook game. This paper is complementary to ongoing work [7] examining 'speed' as a perceptual intermediate used to model higher level semantic attributes such as sadness vs. happiness.

In this work we address the following questions:

- Can we use a GWAP for collecting tapped tempo data, and if so how?
- Is the distribution of tempi of each musical example a good predictor of speed?
- If we let users tap rhythms freely along to music, can we find relevant patterns in the data?

The remainder of this paper is organised as follows: Section 2 describes the software architecture and the design of the two mini-games used in this study. Section 3 presents the collected results. Section 4 discusses the data analysis and reflects on the method. Section 5 summarises the results and discusses future work.

## 2. METHOD

In this section, we describe the application architecture as well as the design for the tempo tapping and rhythm tapping experiments with GWAPs.

### 2.1 Application architecture

The GWAP presented here is built with the CaSimIR API and game framework. CaSimIR as well as the method and the user interface to collect similarity data have been introduced in [5].

[1] http://apps.facebook.com/spottheoddsongout/ and http://mi.soi.city.ac.uk/camir/game/
[2] https://www.spotify.com/se/about-us/press/information/
[3] http://www.apple.com/pr/library/2010/02/25iTunes-Store-Tops-10-Billion-Songs-Sold.html

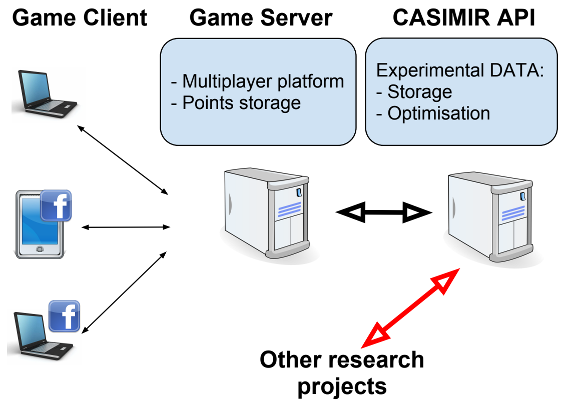


**Figure 1**. The application is organised in three layers: the client, the CaSimIR game server and the CaSimIR API. The API deals with the collection, organisation and selection of experimental data. The game server is separated from the game client to make the multi-player game accessible across different devices and operating systems.

The CaSimIR framework supports the development of social games with a purpose, providing a multi-player platform, high scores, social network integration and compatibility over a large range of browsers and devices. The CaSimIR API is a machine-to-machine interface between the game and the database systems, providing storage and controlled selection of experimental data.

From the developer's point of view, the client interface structures the game as a sequence of mini-games, which are part of the modular architecture. In the latest release of the Spot The Odd Song Out game, we provide three mini games studying different aspects of music: music similarity, tempo and rhythm.

#### *2.1.1 The CaSimIR API: the data collecting system*

The CaSimIR API provides an interface between the game and the database. Its two main purposes are firstly to gather and relate the data from different instances and different applications into a central database. Secondly the API also manages the selection of stimuli in order to achieve intended data properties, e.g. a certain number of subject responses per stimulus, or connectedness of graphs in the result coverage.

The API controls the number of responses for each song and returns a song according to the intended data set properties. In the "tap tempo" mini-game, for example, the API checks whether 70% of the songs have been annotated at least 7 times. Once this condition is achieved, new songs will be added to a subset and presented to future users.

Each time a player joins the game, they are authenticated in the API according to a unique key related to their IP or Facebook profile. Similarly, songs are uniquely referenced from the MagnaTagATune [1] and the Million Song Dataset [8] dataset. Thus data can be related by song and user across different games, supporting the comparison and aggregation of results from different studies.

#### *2.1.2 The CaSimIR game framework*

In comparison to traditional experiments or to web surveys a GWAP has additional requirements. A game design needs more software functionality to provide an engaging experience. Especially the cooperative aspect of a multi-player increases enjoyability and involvement of the subjects, but it poses further challenges: e.g. players with variable latencies, different Java Script interpretation across browsers and the need for AI-players to avoid empty matches. CaSimIR aims to provide a modular multi-player game environment that many projects can easily adapt to their needs, without having to re-implement basic functionalities such as data management, player synchronisation or social sharing.

Most existing GWAPs simulate a multi-player experience or restrict interaction to a high score table. In contrast, Spot The Odd Song Out features almost real-time interaction, a display menu, high scores and Facebook integration in addition to the mini-games for data collection. All mini-games feature basic gaming functionality such as a navigation menu, volume control and the display of the status of collaborating players.

To encourage players to return, options to customise the game experience are provided: Players may use points earned before to buy a new avatar or a genre in the music similarity mini-game. We also provide high scores tables, modern graphics, and social advertisement on Facebook to attract players.

#### *2.1.3 Client-side JavaScript and implementation issues*

The game client runs on mobile devices and computers in a web browser supporting HTML5. We use LimeJS [4] game-framework and the Google Closure Library [5] to achieve compatibility over many devices and browsers. By providing a tested multi-platform framework, CaSimIR makes it easier for researchers to develop GWAPs.

### 2.2 Game experience and user interface

The user plays sequence of mini-games and against three other players. Each mini-game corresponds to one experiment, the current succession being "odd-one-out", "tap tempo", "odd-one-out", "tap rhythm", "odd-one-out", "tap tempo". The "odd-one-out" mini-game is described in [5]. During each mini-game, the player is asked to perform a task within 60 seconds. Once all players have completed the task, or on time-out, the results are compared and points are awarded.

We aimed to make the games easily understandable with short explanation. In the design stage we found that implicit information from images, titles and overall layout has a stronger influence on the user than lengthy instructions. Thus the tasks are described in few short sentences in the first appearance of each mini-game and descriptive images and animations are provided. In the first run of a mini-game, the interface provides additional information.

[4] `http://www.limejs.com/`
[5] `https://developers.google.com/closure/library/`

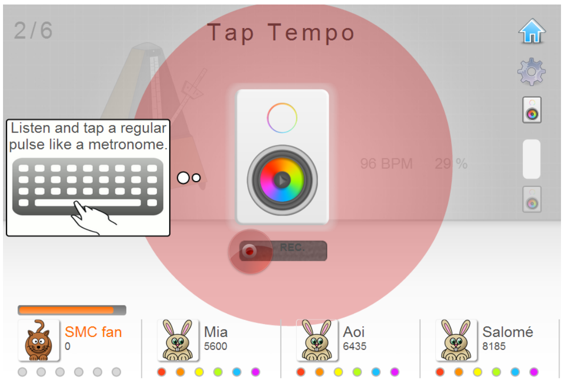


**Figure 2**. Screenshot of the "tap tempo" mini-game.


**Figure 3**. Screenshot of the "tap rhythm" mini-game.

The data of these runs is still recorded but can be identified in evaluations.

We use rules and rewards to encourage "well-behaved" responses, and avoid cheating or random behaviour. It is also important not to bias the experiment by rewarding very particular inputs. We use two approaches for awarding points: basing rewards on a parameter that is independent of the studied parameter (all the tempo octaves correspond to a correct answer) and rewarding agreements of players.

The dataset for the experiments described here contains audio for 100 ring tones synthesised from MIDI and songs from the Million Song Dataset [8] and is used for both the "tap tempo" and the "tap rhythm" mini-games.

### 2.2.1 *The "tap tempo" mini-game*

The "tap tempo" mini-game is designed to study how players tap a tempo. As the mini-game appears, an instruction explains the task. It shows an animated icon of a finger hitting the space bar of a keyboard and a note: "Listen and tap a regular pulse like a metronome." A large icon of a metronome in shown the background. The user listens to the audio clip while clicking on the mouse or hitting any key of the keyboard to reproduce the perceived pulse. Depending on the speed of the tapping, the timings of 8 to 16 taps performed by the player are recorded in ms. At each tap a red flash provides visual feedback. The player has to wait for the other players to finish the task before being shown the results and the rewards. The tempo and the relative precision error accumulated during tapping are displayed in the evaluation.

For the evaluation we only use the intervals between the taps, because the tap positions in relation to the music are subject to latencies that we can not control. The four players are ranked and get 0, 5, 10 or 20 points, based on the ranking score $R_{tempo}$, where lower values are better. $R_{tempo}$ is the weighted sum of the irregularity indicator $ind_{reg}$ and the imprecision indicator $ind_{pre}$

$$R_{tempo} = c_1 ind_{reg} + c_2 ind_{pres} \quad (1)$$

We manually determined $c_1 = 0.1$ and $c_2 = 0.1$ so that the reward is low for users with incoherent balance between regularity and precision.

Let vector $t$ contain the times of the different taps, $T$ the median time difference of successive taps. We define the irregularity indicator

$$ind_{reg} = \sum_{i}^{n} \frac{t_i - t_{i-1}}{T}. \quad (2)$$

The imprecision indicator needs to be minimal for octaves of the tempo, thus we use

$$m = \max\left(\frac{T}{T_{ref}}, \frac{T_{ref}}{T}\right) \quad (3)$$

$$ind_{pre} = m - round(m) \quad (4)$$

where $T_{ref}$ the reciprocal of the tempo. $ind_{pre}$ will also be minimal for integer multiples greater than 2, 3 authorise a ternary subdivision of the bar and higher values did not appear in our data.

In the results screen following each mini-game, the tempi given by all the players are shown coloured from green to red depending on the imprecision indicator. The relative error in percent is displayed in the same way according to the irregularity indicator. The earned points (0 - 20) are also displayed.

### 2.2.2 *The "tap rhythm" mini-game*

The "tap rhythm" game is designed to encourage complex rhythmic performances. The screen shows four circles of different colours and a drum kit is shown in the background. On each of the circles we display the letter D,C,J or N and a picture of a djembe and a double bass. By hitting the keys for the letters users can tap different rhythms or instruments which together form a rhythmic pattern. An instruction bubble explains: "Reproduce the main rhythmic pattern. Repeat it during 9 sec. Use four fingers." When the user hits one of the four keys, the corresponding circle blinks as a visual feedback. After pressing play and taping for the first time, the user's taps are recorded during nine seconds. When the rating is displayed an instruction bubble explained that it is based on complexity and precision of the tapping sequence. The users are ranked according to their performances and get from 0 to 20 points. In the rating computation and the presented preliminary analysis, the recorded taps of the four keys are merged to a single tap sequence.

We define the precision indicator on the same way as in the "tap tempo" mini-game, but the inverse of the frequency maximising the squared spectrum of the tap sequence is used instead of the median time difference of successive taps.

The complexity indicator ranges from 0 to 36 and is defined as growing each time one of the three highest peaks of the squared spectrum goes below one of the threshold values 0.025,0.02,0.016 and 0.013. We manually determined these threshold values so that:

- a random performance will obtain the three peaks under the thresholds values,
- an isochronous sequence will have the second and third peaks lower than the thresholds,
- poly-rhythms - repeating a one bar pattern containing multiple intervals - are promoted by giving three peaks over the thresholds.

The players are ranked according to the value:

$$R_{rhythm} = \frac{1}{1 + d_1 ind_{comp}} + \frac{1}{1 + d_2 ind_{pres}} \quad (5)$$

The constants $d_1 = 0.05$ and $d_2 = 0.2$ are determined manually like the threshold values.

These rating functions are not optimised estimators of tempo, rhythmic accuracy or complexity, but we feel that they are meaningful enough to support enjoyable game play and encourage participants to enter meaningful data.

## 3. PRELIMINARY RESULTS

The presented results are based on data collected during an internal testing period of two weeks and one week following the official release on Facebook and the web. The database contains 904 tempo estimations and 396 rhythm estimations. These were provided by 114 Facebook users and 50 further unique users of the web version. For the majority of these users, attributes including age, gender, country and further demographic data have been collected. A measurement of the accuracy of the recording of the taps led to an error below 30 ms in most cases and 100 ms in one particular case.

### 3.1 Testers' feedback

During the internal testing we provided computers and iPad devices to the players, observed how they understood the tasks and asked them for feedback following each match. Testers found the game was enjoyable but the mini-games would deserve more explanations to be understood from the beginning. Many testers were unsure what to do in the "tap rhythm" mini-game. Many non-musician players expressed that they felt the "tap rhythm" and sometimes the "tap tempo" games were too difficult. However, testers with a musical background appreciated this part of the game.

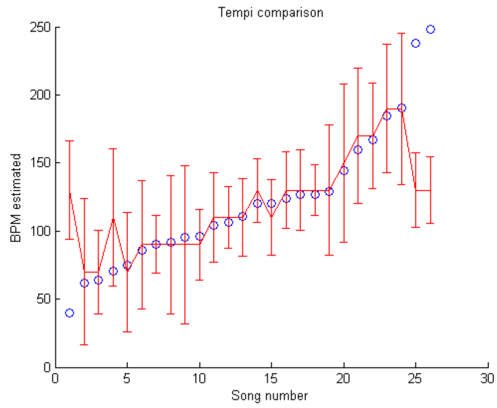


**Figure 4**. Comparison of tapped tempo (red line) vs ground truth (blue dots). The red curve follows the dominant tempo estimated from tapping. Red bars indicate the standard deviation of the user data. The computed/expert ground truth tempo and the tapped tempo agree in most of the cases. Extreme tempi tend to be tapped as multiples or divisions in a range of about 70 to 180 BPM.

### 3.2 Precision of tempo estimations

In order to filter out tap sequences given by players who did not actually try to perform the intended task, we used a set of "relevance thresholds". With the following thresholds for the "tap tempo" we obtain 46% of relevant tap sequences:

- the average tempo is between 30 and 300 BPM
- the relative standard deviation of time intervals is below 25%
- the maximum of relative deviation is under 35%

Players without musical background were sometimes adapting their taps with strong rhythmic changes or paused tapping. Most of those values are filtered out.

We compared the tapped tempo to a computed ground truth tempo: The tempo extraction is based on a percussive onset detection in the audio files [9] and agrees with the expert tempo for every song of the ringtones dataset, which was mainly used in the "tap tempo" mini-game. For the songs from the Million Song Dataset, we used tempo estimations provided by The Echo Nest.

### 3.3 Tempo and perceived speed

By *speed* we mean subjective ratings of "how fast" a piece of music is on a scale from "not at all" (0) to "very much" (see [10] for more detail). The relation between speed perception and tempo is not straightforward. E.g. the perceived tempo for a certain music example is not necessarily the same for all listeners. There are usually different metric levels present at the same time in a piece, and the one that is chosen as the most salient tempo can vary among listeners. This can be referred to as the tempo octave issue studied in [11–13]. For our study, we use three different tempo estimates:

| Method | R-Square |
|---|---|
| Centroid | 0.59 |
| Tapped | 0.51 |
| Expert | 0.61 |

**Table 1**. Linear regression of the correlation between speed and the listed variables.

- expert tempo - estimated by a music expert,
- tapped tempo - the most frequently tapped tempo within the players,
- computed tempo - is determined by an algorithm [9].

Madison et al. [10] relate expert tempo to speed ratings, concluding that speed can be modelled as a sigmoid function of an expert tempo. Elowsson et al. [9] find that a computed tempo and a combination of custom features for onset and computed tempo could predict up to 90% of estimations.

Levy [11] used a web based application to collect the speed labelled as fast, intermediate or slow while asking the user to tap the tempo. The purpose was to correct possible tempo octave errors in a computed tempo. Determining speed from the tempo distribution with this dataset is not straightforward, because only three categories of speed are used and a bias may be introduced by the subject being asked to tap and evaluate speed at the same time.

### 3.4 Tempo distribution

For each song we computed the centroid of the tempo distribution given by the tapped estimations. Madison et al. [10] relate expert tempo values to speed estimations with a sigmoid curve. We computed a linear regression to compare the correlation between perceived speed collected in a separate experiment described here [7] and the centroid of the tapped distribution, the tapped tempo, and the tempo ground truth. The results are summarised in a table 1 and figure 5.

### 3.5 Rhythmic pattern identification

The data acquired from the second experiment are analysed as an onset list. Based on this list we compute a main rhythmic pattern description: To each onset we associate a Gaussian function with a standard deviation of 50ms. We define the bar period as four beats in the ground truth tempo. Inspired by the Beat Spectrum published by Jonathan Foote et al. [14], for each offset we sum the corresponding positions of the tap signal over the bar periods starting. This results in a pattern representing the accumulated tap incidence over the time of one bar. We sum over all performances to obtain an estimation of the main tapped pattern of the song. This pattern is compared to a pattern extracted using a computed onset list over the audio file. In this experiment the offset is hard to define, as for recorded taps, the time when audio playback starts could

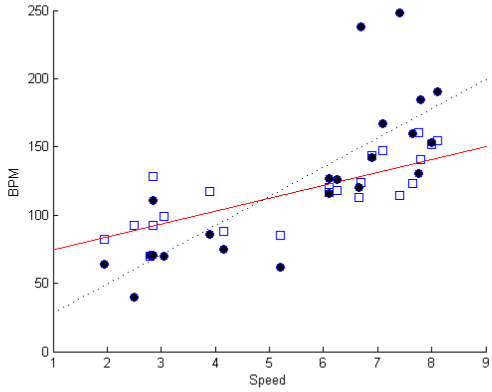


**Figure 5**. Regression for tapped tempo centroids (squares) and ground truth (filled circles). Regression of centroids is plotted as red line (75 - 200BPM), and regression of tempo ground truth is plotted as black dotted line. The X-axis represent the speed and the Y-axis represent the tempo.

not be accurately recorded. We recreated the offset by assuming that the player taps on the first beat more than at any other time in the bar.

## 4. DISCUSSION

### 4.1 General assumption on tempo estimations

In nearly all cases (figure 4), the tapped tempo data reproduced the ground truth tempo given by computer and experts, with tempo octave disagreements occurring only for extreme tempi. When a tempo octave can be ambiguous, indicated by more than one valid tapped tempo centroid, non-musician players disagreeing with the ground truth tend to prefer tempi close to 120 BPM.

### 4.2 Tempo distribution

The centroids of the tempo distribution and the expert tempi in figure 5 show similar performances as indicators of the speed ratings. Madison et al. [10] indeed describe a high correlation between speed and expert tempo. We have not yet collected sufficient data to allow for a more wide-ranging comparison. The above correlations of tempo centroid and expert tempo are still encouraging to infer perceptual speed from tempo centroids. The tapped tempo itself proved an inferior predictor of perceptual speed.

For musically trained subjects, tempo-related tasks appear relatively easy. On the other hand, non-musician subjects had often great difficulty in reproducing the tempo. MacDougall et al. [15] state that 120 BPM correspond to a "resonant frequency" of the human body. Madison et al. suggest the tendency of performing a tempo in a middle range tempo between 90 and 150 BPM. A non-musician may have trouble to produce a pulse out of this natural range from a motor point of view. On the other hand, a music expert will be able to pick a tempo multiple more representative of his perception of speed in the musical piece. This would explain both the disagreement of the crowd and

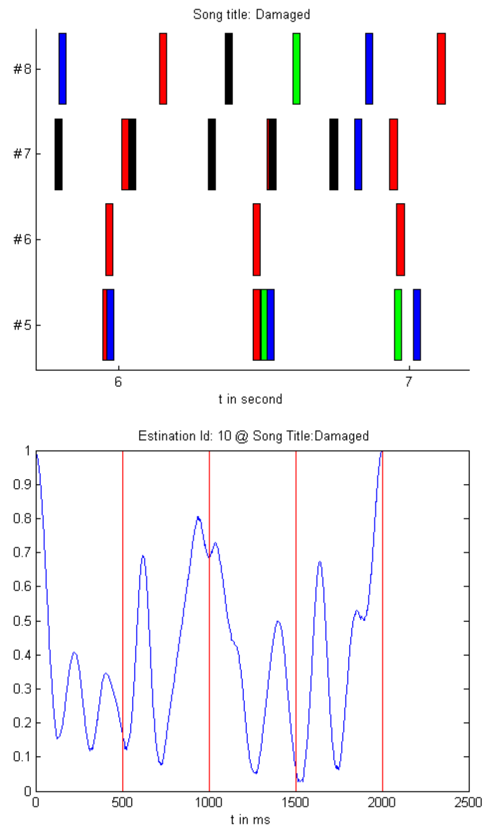


**Figure 6**. (top) Rows depict onset times from “tap rhythm” performances for 4 users, with colours identifying different fingers. (bottom) Accumulated tap incidence pattern for a single performance, red vertical lines locate predicted beat positions.

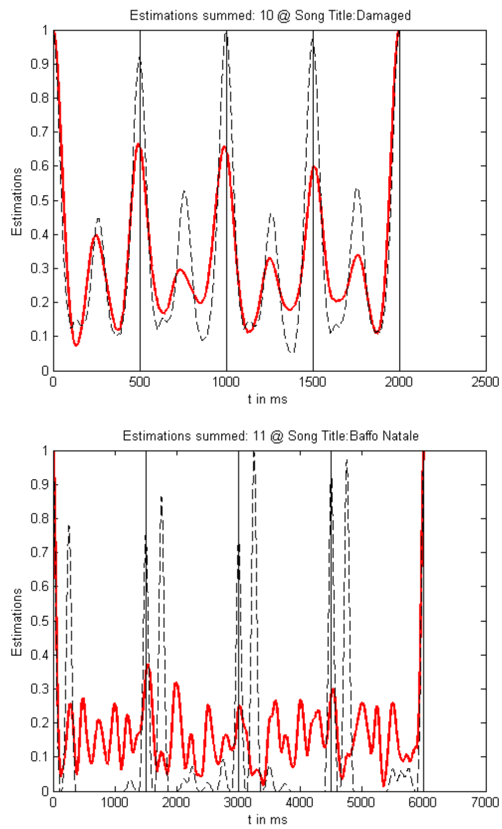


**Figure 7**. Averaged tap incidence pattern over several performances for 2 songs (top) Song “Damaged”, time signature 4/4, averaged over 10 data entries. (bottom) Song “Baffo Natale”, time signature 12/8, averaged over 11 data entries. The beat patterns are plotted as continuous graphs (-). For comparison, the dashed curves (–) represent beat patterns automatically extracted from audio.

the expert in extreme ranges as well as the higher correlation for the expert tempo compared to the most frequently tapped tempo.

### 4.3 Tapping free rhythms with four fingers

Many different behaviours can be identified in the collected data (Figure 6). The task was designed to give interpretive freedom to the player, in contrast to the very strict “tap tempo” mini-games. Unfortunately, many players perceived it as hard to understand and perform. This may be due to the interpretative freedom and multi-limb and finger coordination as well as rhythmic skills required by the task. The main aim of this second task was to identify which metric positions would be emphasised by players. This could have led to more complex or irregular patterns such as syncopation, swing or groove, useful for extending the regular notion of tempo.

During analysis we found that, by summing the main pattern tapped by ten players or more, (see Figure 7) we converge to a hierarchical description of the time division in the bar. In most of the cases, this hierarchical description corresponds to the actual time signature a music expert would assign to the music piece: a hierarchical sequence of regular subdivisions of the bar is indicated by the most represented onsets. With a standard $4/4$ (four quarters) audio stimulus, a clear majority tapped beat corresponds to the first beat of the period. However, there is a bias here, as we do not know the absolute time of each tap. To overcome this problem we shift each performance’s maximum of accumulated tap incidence to the beginning of the pattern. The lower peaks represent the three other beats of the bar. Eights are tapped as well, but no lower subdivisions of the bar are represented.

Figure 7 compares players’ patterns to the patterns extracted from the audio clip. In most of the cases, the tapped pattern corresponds very closely to the onset identified in the song. Yet in a song containing twelve subdivisions per bar, players agree on the song signature even if the automatically extracted onsets barely finds this in the music.

In this preliminary study, we have considered the tapping with the four fingers as having an equivalent role. This representation led to an interesting result but did not represent rhythmic irregularity of some songs as expected. The dataset could be further explored, e.g. by comparing players using 1,2,3 and 4 fingers or identifying reproduced pattern in a particular bar of the song.

## 5. CONCLUSIONS AND FUTURE WORK

In this paper we have described the CaSimIR API, a modular GWAP framework, and evaluated its applicability to research in tempo, rhythm and speed using two experiments: "tap tempo" and "tap rhythm". The CaSimIR framework allowed for the recording of taps in reasonable accuracy. The programming of automated answer scripts via bots should be discouragingly complex. This is due to integrity testing of submitted user data and the general complexity of the interaction with the game-style user interface.

### 5.1 Results of the tapping experiments

For "tap tempo", we were able to collect a large dataset of new tempo estimations, and found the collected data strongly reproduced ground truth data, encouraging the use of GWAP's to collect more tempo information about music, which could be used for inferring perceptual speed.

The "tap rhythm" mini-game also allowed for the collection of rhythmic patterns. Although we encountered some usability problems, rhythmic patterns were extracted and used for analysis by combining data from several users. When asking a group of people to perform a rhythm freely with four fingers, the data was relatively noisy. However, the averaged pattern still converged to a regular hierarchical subdivision of the bar similar to a traditional time signature.

After a preliminary data collection period of a few of weeks, we raise questions to be explored in future studies: In order to further investigate the relation of speed perception and collected tempo data, we need precise numerical results and more data. The exploration of the "tap rhythm" dataset encourages further research into characterising rhythm singularities or particularities in motion related to the reproduction of rhythms. Results of these experiments might also relate to other perceptual features such as rhythmic complexity or clarity.

### 5.2 Lessons learned in using a GWAP

The CaSimIR framework addresses many challenges related to the GWAP approach: It provides a multi-player platform, survey example selection, and manages data storage allowing for combination of different data collection ventures.

The game has reached enough players to observe interesting details using this experiment. Although the mass effect of a GWAP based experiment is appealing, it is not granted: Many players did not return to the game, and it proved hard to maintain a constant amount of players participating. Other GWAPs such as HerdIt [2,3] have reached a threshold of 500 players, TagATune even reached an audience of almost 15,000 players. Reaching such a success is a hard task and requires appealing game as well as a long term effort including regular additions and advertising. The number of players needed for particular studies may not justify the complexity of the application compared to a study with large promotion [11].

As earlier games focussed on collecting textual annotations, we show that recent technology allows for collecting tempo data on a large scale of users. The tightly timed game interaction promotes high attention of users, but means of controlling the users' context such as noise, type of speakers and sound levels could not be applied. Although we found the visual feedback helpful, we did not measure the influence on the performer. Despite the preliminary nature of our data collection and the moderate timing precision, our data still allowed to validate and identify human perception specificities, which is appropriate for a preliminary study. Communication and interaction between users is important and should be improved, as collaborative playing might be a key to attract more players.

## 6. REFERENCES


[1] E. Law and L. von Ahn, "Input-agreement: a new mechanism for collecting data using human computation games," in *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, 2009.

[2] D. Turnbull, R. Liu, L.Barrington, and G.Lanckriet, "A game-based approach for collecting semantic annotations of music," in *Austrian Computer Society*, 2007.

[3] L. Barrington, D. Turnbull, D. O'Malley, and G. Lanckriet, "User-centered design of a social game to tag music," in *Proc. of the ACM SIGKDD Workshop on Human Computation*, 2007.

[4] M. L. Mandel and D. P. Ellis, "A web-based game for collecting music metadata," *Journal of New Music Research*, 2008.

[5] D. Wolff, G. Bellec, and T. Weyde, "A music similarity game prototype using the CASimIR API," in *Online Proc. of ISMIR Late Breaking Sessions*, 2012.

[6] L. von Ahn, "Game with a purpose," *Computer*, 2006.

[7] A. Elowsson, A. Friberg, and G. Madison, "Tempo estimation of popular music," in *Submitted*, 2013.

[8] T.Bertin-Mahieux, D. Ellis, B.Whitman, and P. Lamere, "The million song dataset," in *Proc. of the 12th International Society for Music Information Retrieval Conference*, 2011.

[9] A.Elowsson and A.Friberg, "Modelling perception of speed in music audio," in *Forthcoming for Proc. of SMC*, 2013.

[10] G. Madison and J. Paulin, "Ratings of speed in real music as a function of both original and manipulated beat tempo," *Acoustical Society of America*, 2009.

[11] M. Levy, "Improving perceptual tempo estimation with crowd-sourced annotations," in *Proc. of ISMIR*, 2010.

[12] M. McKinney and D. Moelants, "Deviations from the resonance theory of tempo induction," in *Proc of Interdisciplinary Musicology*, R. Parncutt, A. Kessler, and F. Zimmer, Eds., 2004.

[13] J. Hockman and I. Fujinaga, "Fast vs slow: Learning tempo octaves from user data," in *Proc. of ISMIR*, 2010.

[14] J. Foote, "The beat spectrum: a new approach to rhythm analysis," in *In Proc. IEEE Int. Conf. on Multimedia and Expo*, 2001.

[15] H. G. MacDougall and S. T. Moore, "Marching to the beat of the same drummer: the spontaneous tempo of human locomotion," *Human Movement Science*, 2005.