Visualizing Air Drums: Analysis of Motion and Vocalization Data Related to Playing Imaginary Drums

Ana Julia Pereira Caetano[1 2] **, Tiago Fernandes Tavares**[1 2]

[1]School of Electric and Computer Engineering (FEEC) – University of Campinas (Unicamp)
CEP 13083-852 – Av. Albert Einstein, 400 – Campinas – SP – Brazil

[2]Interdisciplinary Nucleus for Sound Studies (NICS) – University of Campinas (Unicamp)
CEP 13083-872 – Rua da Reitoria, 165 – Campinas – SP – Brazil

tavares@dca.fee.unicamp.br

Abstract. *Air drums, or imaginary drums, are commonly played as a form of participating in musical experiences. The gestures derived from playing air drums can be acquired using accelerometers and then mapped into sound control responses. Commonly, the mapping process relies on a peak-picking procedure that maps local maxima or minima to sound triggers. In this work, we analyzed accelerometer and audio data comprising the motion of subjects playing air drums while vocalizing their expected results. Our qualitative analysis revealed that each subject produced a different relationship between their motion and the vocalization. This suggests that using a fixed peak-picking procedure can be unreliable when designing accelerometer-controlled drum instruments. Moreover, user-specific personalization can be an important feature in this type of virtual instrument. This poses a new challenge for this field, which consists of quickly personalizing virtual drum interactions. We made our dataset available to foster future work in this subject.*

1 Introduction

Gesture-controlled virtual instruments can provide musicians an experience closer to that provided by acoustic instruments. This experience relies on gesture acquisition and instrument emulation [1]. However, not all acquisition or emulation methods can lead to musically meaningful instruments [2].

Musical meaningfulness can be pursued by a design process involving emulation of real, physical environments. For such, it is possible to use pattern recognition techniques. These techniques can be used to detect specific gestures, as well as their intensity and possible variations, and link them to sonic manifestations [3]. However, they require a reasonable amount of labeled data for parameter optimization [4].

Gesture-related labeled data is hard to obtain because it needs to be acquired from human subjects. Human acquired gesture data can account for gesture variations that are hard to predict with physical motion models. In addition, humans have particular prior experience and expectations regarding the behavior of virtual instruments [1].

This phenomenon has been studied by Maki-Patola [5], who designed an experiment in which subjects played virtual drum instruments using different interfaces. This experiment showed that playing precision varies according to people and interfaces. Maki-Patola used fixed tempo and predefined interfaces and interactions to emulate real acoustic drums.

Another approach to problem of emulating virtual drums was presented by Havel and Desainte-Catherine [6]. They proposed a virtual drum instrument specially designed for a specific musician. This instrument provides an interaction model that involves strike classification in addition to the detection. However, all collected data is related to one subject, therefore it cannot be generalized.

Another initiative towards the analysis of percussive gestures was performed by Dahl [7]. This study was based on a dataset consisted of free-hand movements acquired while subjects tried to synchronize to a pre-recorded rhythm. Dahl [7] studied the position, velocity and acceleration of the subjects' wrist and hand movements.

In this work, we present a dataset containing gestures data collected from 32 different subjects playing imaginary drums without accompanying music, which consists of a different condition that that analyzed by Dahl [7]. The dataset also contains vocalizations of the expected sonic results for each subject. The data acquisition process did not induce subjects to play in a particular *tempo*. Our dataset can be used for the construction of machine-learning based instruments that generalizes across different people.

We also performed data analysis showing that the alignment between the vocalization and its gesture signal is different for each subject. This difference can be observed regardless of their previous musical experience and rhythmic intention.

The remainder of this work is organized as follows. Section 2 describes the data acquisition process. Section 3 presents further analysis on the acquired data. Last, Section 5 concludes the paper.

2 Data Acquisition

Our data acquisition process relies on the assumption that different people expect different sound results when they play imaginary drums. With that in mind, we designed a data acquisition process in which subjects provide both gesture data and its respective expected vocalization.

The dataset contains data acquired from 32 sub-

jects (23 male and 9 female), aged between 17 and 65 years old. Within this group, 20 subjects had previous musical experience and 12 did not (10 of them had experience in playing percursion, while 10 did not). All subjects are residents of the South-East of Brazil. They all signed a free consent form. This experiment was approved by the Ethics Committee of the University of Campinas (CAAE 53738316.0.0000.5404). Each subject was instructed to perform gestures that emulate playing an imaginary drum using a WiiMote as a stick, and vocalize the sound they imagine to produce. Each subject recorded two different tracks. In one of them, the subject was instructed to maintain a steady rhythm and *tempo*. In the other, they were instructed to perform free beat variations.

Audio was acquired using a laptop microphone and the WiiMote device data was upsampled to 44100 Hz. As a result, we generated 64 tracks containing time aligned gesture and vocalization. On average, each track is 8 seconds long. We only used the X axis of the WiiMote accelerometer because the acquired motions are more closely aligned to this axis.

In the next section we conduct further discussion about observed data.

3 Data Analysis

Our data analysis was based on observing the alignment between gestures and their corresponding vocalizations, as shown in Figure 1. As will be further discussed, this alignment varies, which indicates that different subjects imagined different interactions with their imaginary drums. We also observed the percussive gestures shape variations across subjects and acquisition conditions.

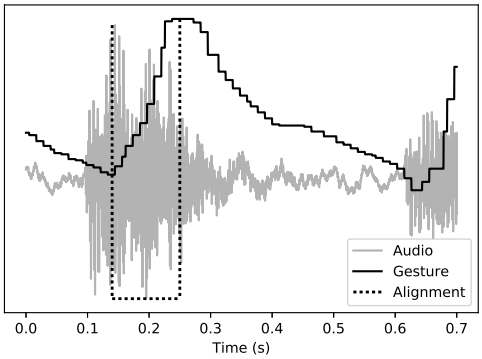

Figure 1: Audio and gesture alignment. It is possible to observe that the peak gesture activity happened after the percussive vocalization. This behavior was not consistent among the acquisitions.

Differences were observed regardless of the subjects' previous experience with percussive instruments. This aspect is further discussed in section 3.1.

We also observed that the relationship between gesture and its imagined sound changes for the same person according to their rhythmic intention. This means that performing different rhythms impacts on this relationship. A deeper discussion on this subject is conducted in section 3.2.

We selected data from specific subjects as shown in Table 1. The same subjects were used in the analyzes conducted in sections 3.1 and 3.2.

Table 1: Subjects selected for data analysis.

Subject	Experience
S1	No musical experience
S2	Non-percussive instrument experience
S3	Percussive instrument experience
S4	Percussive instrument experience
S5	Percussive instrument experience
S6	Percussive instrument experience

3.1 Impact of Previous Percussion Experience

Figures 2 and 3 show audio and gesture captured from subjects with no previous experience in percussion. Data shown in Figure 3 relates to a subject with experience in non percussion instruments. It is clear that the peaks and valleys related to the performed gesture and vocalization align differently for each subject. Moreover, the musically inexperienced subject (S1) presents less consistency in this alignment than the experienced one (S2).

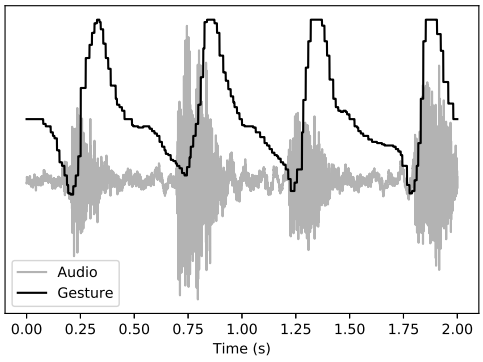

Figure 2: Audio and gesture related to a steady rhythm performed by S1.

Figure 4 shows the data acquired from S2, a subject with previous experience in percussion instruments. It is possible to observe that gesture and vocalization are aligned at their onsets and a valley in the gesture preceeds the vocalization. Figure 3 shows that this alignment can also be observed for S3. However, it is possible to see that S3 produces a sequence of two valleys before the vocalization, and two peaks after the vocalization, while S2 produces a single valley and a much smaller second peak. Also, the vocalization of S3 is closer to its preceding valley when compared to the vocalization of S2. This suggests that the imagined interaction is different for each one of them.

Figures 5, 6 and 7 depict the data captured from three other percussionists (respectively, S4, S5, and S6)

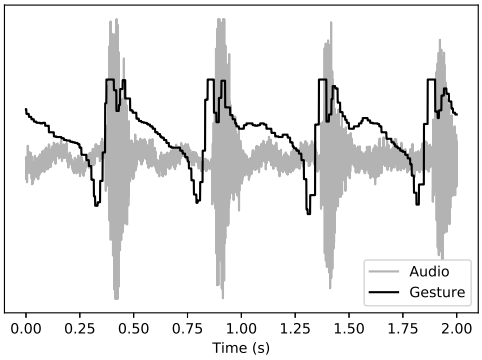

Figure 3: Audio and gesture related to a steady rhythm performed by S2.

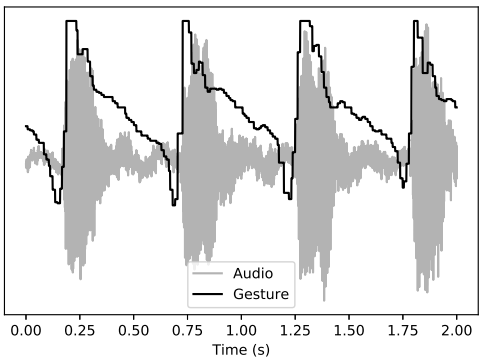

Figure 4: Audio and gesture related to a steady rhythm performed by S3.

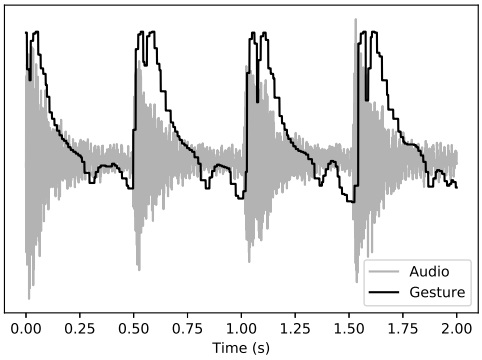

Figure 5: Audio and gesture related to a steady rhythm performed by S4.

Interestingly, data from S5 also shows alignment between the vocalization and the gesture activity valley, but this cannot be observed in S6 or S3. Also, the alignment

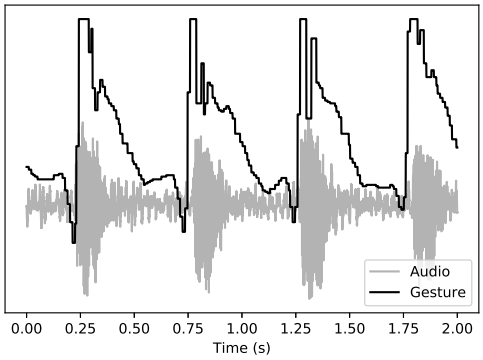

Figure 6: Audio and gesture related to a steady rhythm performed by S5.

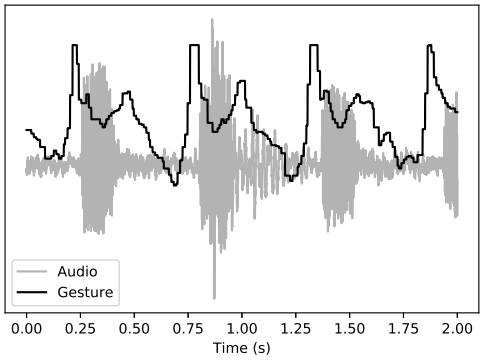

Figure 7: Audio and gesture related to a steady rhythm performed by S6.

with different levels of expertise. It is clear that the alignment of motion peaks and valleys with the vocalization is different. This difference is similar to that found between S2 and S3, which indicates that these differences are due to imagining different interactions or situations, not to differences in musical expertise.

between the motion signal valleys and peaks and vocalization data seems to be consistent for each subject. All these data suggest that different people imagine different interactions with the virtual instrument regardless of their previous musical experience.

3.2 Impact of Rhythm

When playing in a varying rhythm, subjects expressed less confidence during data acquisition. We speculate that this is linked to the fact that most people are more used to playing and listening to music with a steady rhythm. Moreover, performing an unknown rhythm in an unknown instrument generated discomfort.

Figures 8, 9, 10, 11, 12 and 13 show data acquired from the same subjects discussed in Section 3.1.

The subject in Figure 8 did not show consistency in the alignment of voice and gesture. This behavior replicates the observation in Figure 2. Moreover, the experienced musicians, as shown in figures 9, 10, 11, 12 and 13, stopped presenting alignment consistency.

Therefore, in all cases is possible to observe inconsistency in the alignment between gesture and vocalization. Also, the shape of their gesture signal varied more within the same acquisition.

Data suggests that subjects did not have a clear

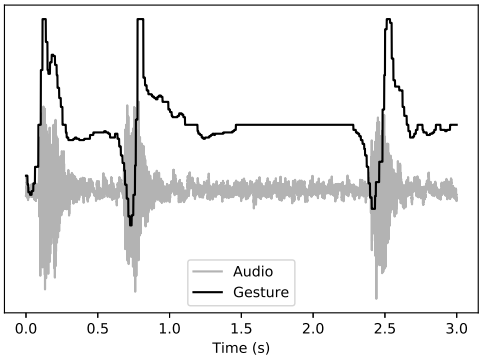

Figure 8: Audio and gesture related to a variable rhythm performed by S1.

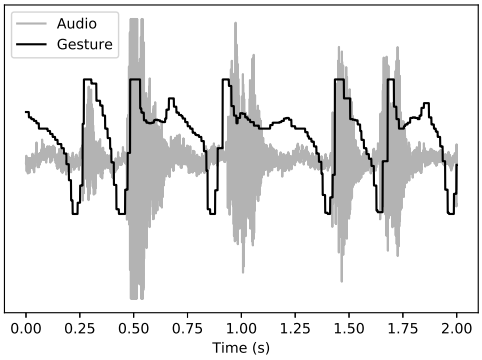

Figure 9: Audio and gesture related to a variable rhythm performed by S2.

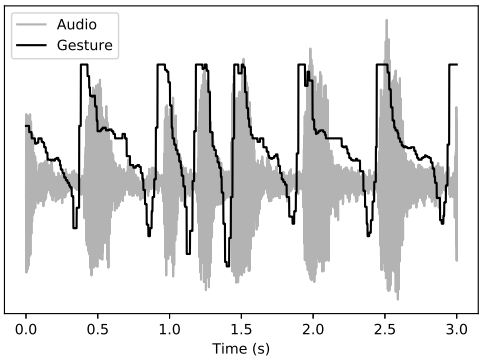

Figure 10: Audio and gesture related to a variable rhythm performed by S3.

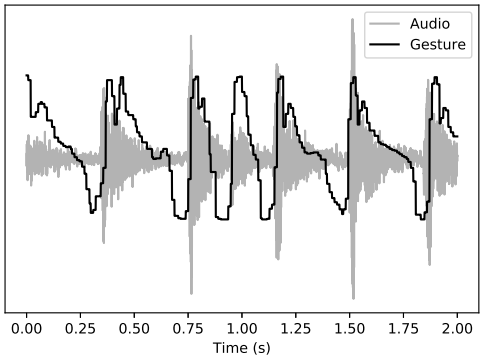

Figure 11: Audio and gesture related to a variable rhythm performed by S4.

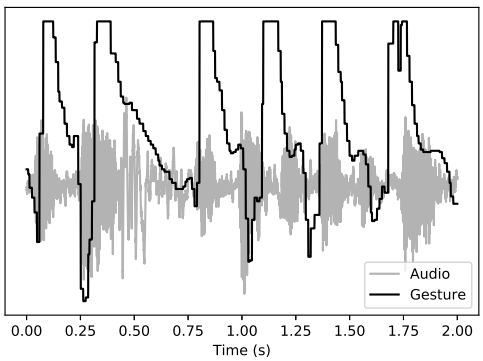

Figure 12: Audio and gesture related to a variable rhythm performed by S5.

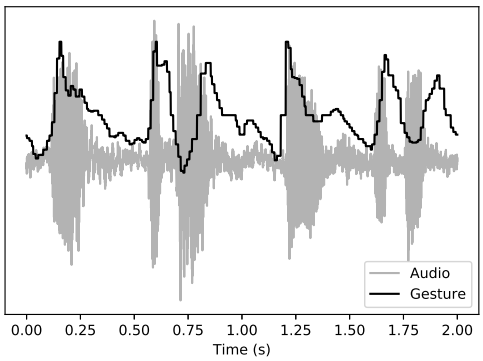

Figure 13: Audio and gesture related to a variable rhythm performed by S6.

idea of how to interact with the virtual instrument without the support of a predefined rhythm. This can be linked to their lack of experience with this specific instrument (imaginary vocalized drums) and these specific conditions, regardless of their general experience with music.

4 Discussion

There are two important aspects that must be noted in the acquired air drum gestures. First, we note that subsequent gestures performed by the same user tend to be similar. Second, we note that gestures performed by different users tend to be different.

The alignment analysis can be performed using the vocalization and the motion signal peaks as references. It is possible to see that S6 performs motions that peak around 0.1 s before the vocalization peak, as shown in Figure 7, while S5 (Figure 5) aligns motion and vocalization peaks and S1 (Figure 2) performs the peak around 0.2 s after the vocalization. This means that the peak can have a difference of up to 300 ms in the alignment due to the

change in the subject. Such a difference is harmful for drum performances.

Similarly, the valley in the motion signal can happen together with the vocalization (S1), around 0.1 s before it (S2), immediately before the vocalization (S5) or up to 0.2 s before the vocalization (S6). This is means that this inter-subject difference is around 200 ms, which is also harmful for drum performances.

It is important to remember that real drums provide both audio and physical feedback. Moreover, the physical feedback strongly correlates to the audio feedback, both in their time alignment and their percussive, "point" quality. As a consequence, it is possible to learn and adapt oneself to the playing of a drum.

On the contrary, air drums are played solely using muscle memory and one's perspective. Even if virtual drums can yield audio feedback, they cannot provide physical interactions. Hence, the playing differences are hard to overcome by practicing.

For this reason, user-specific personalization is an important feature for virtual, accelerometer-controlled drums. This is a seldom explored problem in the field of digital musical instruments. In order to foster this type of research, we made our dataset available online at `http://timba.nics.unicamp.br/mir_datasets/gesture/wiimote_ajpc.zip`.

5 Conclusion

In this work we built a dataset containing both gestures and vocalizations related to a virtual percussion instrument imagined by subjects. This dataset is available at `http://timba.nics.unicamp.br/mir_datasets/gesture/wiimote_ajpc.zip`. We analyzed the shapes of the motion signals and their alignment to the corresponding vocalizations.

Qualitative analysis revealed that different persons use diverse motions to play imaginary drums, which corroborates with the observations of Maki-Patola [5]. Also, we see a high inter-subject difference between the alignment of peaks and valleys of the motion signal to their vocalizations is severely different. Nevertheless, we can see a greater intra-subject similarity between gestures when playing steady rhythms, but this similarity decreases when non-steady rhythms are played.

This means that predicting the vocalization beats is a task that requires user-specific personalization. Such a task wil be tackled in future work.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior - Brasil (CAPES) - Finance Code 001.

References

[1] K. Okada, F. Ishizawa, A. Kobayashi, A. Yoshii, M. Sakamoto, and T. Nakajima. Virtual drum: Ubiquitous and playful drum playing. In *Consumer Electronics (GCCE), 2014 IEEE 3rd Global Conference on*, pages 419–421, Oct 2014.

[2] Mike Collicutt, Carmine Casciato, and Marcelo M. Wanderley. From real to virtual : A comparison of input devices for percussion tasks. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 1–6, Pittsburgh, PA, United States, 2009.

[3] Thomas Hermann Tobias Grosshauser, Ulf Grossekathofer. New sensors and pattern recognition techniques for string instruments. In *New Interfaces for Musical Expression*, pages 271–276.

[4] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification (2nd Ed)*. Wiley, 2001.

[5] Teemu Maki-patola. User interface comparison for virtual drums. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 144–147, Vancouver, BC, Canada, 2005.

[6] Christophe Havel and Myriam Desainte-Catherine. Modeling an air percussion for composition and performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 31–34, Hamamatsu, Japan, 2004.

[7] Luke Dahl. Studying the timing of discrete musical air gestures. *Comput. Music J.*, 39(2):47–66, June 2015.