Exploring the spontaneous expression of human finger-tapping

Javier Nistal*, Perfecto Herrera and Sergi Jordà

Music Technology Group
Universitat Pompeu Fabra
08018 Barcelona, Spain
name.surename@upf.edu

Abstract. We present a study on the behavior of human finger-tapping and the spontaneous expression of rhythm. For the purposes of this study we construe interpret finger-tapping as the casual and rhythmic hitting of objects for the expression of music. Our motivation for this study is to connect spontaneous finger-tapping, human-computer interaction and the automatic arrangement of percussion for music creation. Specifically, here we report on the characterization of spontaneous rhythm creation behavior as a prerequisite to develop rhythm-aware music creation interfaces. First, we collect a dataset by recording spontaneous finger-tapping patterns performed by subjects from different music backgrounds. An online survey gathering information about the recording is then submitted to the volunteers. Analysis of the survey answers and low-level audio features suggest that there are two ways for finger-tapping depending on the music skills of the performer (i.e., "experts" versus "naive tappers"). We explore the former hypothesis by conducting a classification task between onsets from both finger-tapping methods. We achieve a 96% of accuracy in recognizing drumming expertise levels (expert vs. naive) by means of using onset-related acoustic features. Results suggest that people with percussion training are more concerned about timbre aspects and, thus, can take advantage of this quality of sound to provide nuances to each stroke when finger-tapping, as opposed to non-expertise individuals.

Keywords: finger-tapping, human-computer interaction, music creation

1 Introduction

Finger-tapping is the casual and rhythmic hitting of objects, which sometimes can be motivated by the expression of a musical idea. Even though it is one of the most straightforward ways for the spontaneous expression of rhythm, few music technologies take advantage of this habit for improving human-computer interaction in the context of music creation. In this regard, we believe that its study represents a first step towards the future implementation of interfaces for rhythm-aware drum arrangement softwares.

* The work reported here is a summary of the MSc thesis that the first author presented in Universitat Pompeu Fabra

Drum beat composition represents one of the most difficult or tedious parts to carry out when composing a music piece. Producers without a wide percussion training or a deep understanding of MIDI (Musical Interface for Digital Instruments) sequencing and editing, may have difficulties for creating a drum pattern in a Digital Audio Workstation (DAW). This is due either to the lack of knowledge from the part of the producer or to the unavailability or limitations of the commercial equipment for interacting with this musical dimension. Examples of such devices are drum machines, MIDI trigger pads or MIDI drums. In general, the basic interface of these tools offer a set of sensors that trigger a certain sound in the computer when they are hit. Many drum machines implement a second layer for creating patterns in real-time (e.g. Roland TR-808, Akai MPC 2000, etc). In other words, there are "pads" that just trigger sounds, but there are also "step sequencers", that trigger sounds in time. Other recent technologies provide with more ergonomic solutions for controlling drums in a computer-based working set. Aero-drums[1] allows to play a drum-kit in the air, without the need of tactile drum-pads, by attaching a set of sensors to the user's limbs. Similarly, the MIDI drum-gloves[2] integrate a sensor in each of the fingers of a glove and sends MIDI messages to a computer. From a more general perspective, Mogees[3] offers a powerful contact microphone as well as an app which allows to train a system for recognizing the acoustic features of any object being stimulated. These features are then used in the synthesis of new sounds, transforming the object into an audio-driven interface for playing any virtual instrument. However, all the former technologies essentially translate the performed sequence of strokes into a MIDI-like symbolic representation. Furthermore, they demand a thorough training and their learning curve is as slow as in any other musical instrument. This fact explains the appearance of new practice apps for helping users improve their finger-tapping skills, such as Melodics[4].

Whereas the above-mentioned technologies are not easy to use, almost everybody is able to tap with their fingers. We can even sense or imagine complex rhythm compositions from a simple and sketchy finger-tapping pattern. In this context, the interest for studying spontaneous finger-tapping emerges naturally. We want to understand not only the acoustical characteristics of finger-tapping but the perceptual and fuzzy rules behind this practice. Research on this topic is very limited and, to our knowledge, few works have oriented the study of finger-tapping from a cognitive perspective. Efforts have been mainly drawn towards the development of query by tapping systems for music retrieval [1,2,3], the perception of tempo [4] and rhythm [5,6], or the diagnose of diseases [7,8]. Few studies have been focused specifically on finger-tapping as an spontaneous behavior. In 1991, a group of musicologists and doctors studied the differences between musically trained and untrained subjects in their ability to follow repetitive rhythmic tonal patterns by finger tapping [9]. They concluded that trained

[1] http://aerodrums.com/aerodrums-product-page/
[2] https://learn.adafruit.com/midi-drum-glove/overview
[3] http://www.mogees.co.uk/
[4] https://melodics.com/

subjects have a more accurate motor timing than the untrained subjects. Following this work, we take a closer look to the behavior of spontaneous finger-tapping and study the needs of naive users in the context of rhythm expression. Our goal is to set some basis for converting this knowledge into a rhythm-aware creative tool for arranging drum patterns.

The remainder of this paper is organized as follows: first, we describe the methodology followed throughout this work. Next, we report on the results of the demographic study and audio experiments and discuss on their possible outcomes. Finally, we enumerate some of the conclusions and suggest directions for future work.

2 Methodology

Our research encompassed three main targets: record a finger-tapping dataset for our experiments, the preparation of a survey to gather demographic information about finger-tapping behavior, and, as a result of the former step, we conducted a finger-tapping expertise classification task using the collected audio recordings generated as part of the survey.

2.1 Audio Collection Acquisition

We recorded a group of 43 western subjects that were asked to perform a finger-tapping pattern using a certain surface. People ranged from 21 to 35 years old and had very different music backgrounds. Volunteers were recruited from the Music Technology Group and Universitat Pompeu Fabra, a few from ESMUC (Escola Superior de Msica de Catalunya) and opportunistically selected people from the street. The only instruction for performing the finger-tapping was to tap repetitively and sequentially a single rhythm pattern decided by the participant using only hands and/or fingers, so as to preserve spontaneity.

We used the following tools during the audio collection acquisition:

– An empty cardboard box that was provided to the volunteers for performing the finger-tapping (Figure 1). This had the purpose of preserving timbre in the audio recordings, allowing the comparison of low-level features.
– A "Yamaha pocketrak PR7" portable stereo recorder, working at 44100 Hz sampling rate.
– Ableton Live DAW for audio editing and tempo annotation. We segmented the raw recordings into excerpts containing a fix number of pattern repetitions starting at the downbeat.

2.2 Finger-tapping Survey

After recording a finger-tapping dataset, we submitted an on-line survey[5] to each of the volunteers that participated in the experiment. This survey gathered general information concerning the respondent's experience with music, the

[5] Survey link

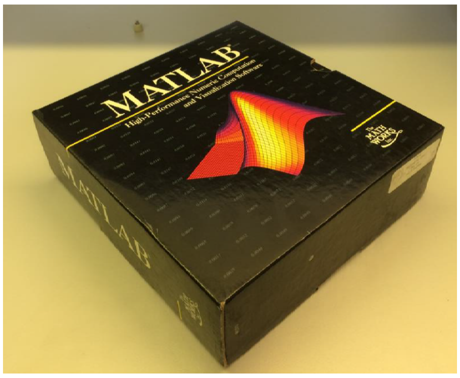


Fig. 1: Matlab box used as surface for finger-tapping

idea behind the performed finger-tapping pattern and the potential capabilities required for a hypothetical expert software for the arrangement of drums. The goal was to understand the musical context of the subjects, what exactly were they thinking when finger-tapping, what they actually performed in comparison to the former, and how would they like their spontaneous tapping to be converted into music. The whole test was designed based on a Likert scale, which is answered using five marks expressing agreement or disagreement[6]. The survey was divided into three parts.

The first part inquired about the musical training of the subjects as well as some demographic details. Special importance was given to their experience in percussion instruments, drum-machines or digital drum manipulation.

The second part of the survey was aimed to gather information about the rhythm performed by the subject and the technique applied for finger-tapping. Our interest focused on:

– The method used for finger-tapping: hands, fingers or both
– The underlying idea behind the pattern: was it performed imitating a real drum-set?, was it based on a music piece?, ...
– The existence of overlapping strokes
– The number of instrument layers that integrated the pattern: is the pattern built of one, two, three or more different instruments?
– The use of the timbre characteristics of the tapped surface: are there intentional timbre nuances between strokes?

2.3 Finger-tapping Expertise Automatic Classification

For this experiment we used a dataset of finger-tapping onsets extracted from previous recordings. The sequence of steps applied to each of the recordings was

[6] Completely agree (CA), Agree (A), Neutral (N), Disagree (D) and Completely Disagree (CD)

as follows: expertise class annotation, onset detection and feature extraction. The annotation was done by manually tagging the performer's expertise class, non-experts (NEP) or experts (EP) in percussion, in each of the finger-tapping excerpts. Likewise, each of the extracted onsets were dumped into a feature dataset containing a field identifying the expertise level of the performer. Features were then used to build a computational model for finger-tapping expertise classification using different machine learning techniques. Evaluation of the models was carried out using N-fold cross validation.

3 Results

3.1 Audio Collection Acquisition

We obtained a set of 43 finger-tapping audio recordings. Each of them were segmented into a fix number of repetitions of the main rhythm pattern. From the 43 subjects associated to each recording, 22 answered the survey that allowed to categorize each finger-taping excerpt. From these sub-set, 14 excerpts were kept for the expertise classification task and the rest had to be dropped due to the level of noise. In table 1 we present a summary of the different sub-datasets used throughout the experiments.

Table 1: Summary of the datasets used for each experiment

	number of items
Survey	22 / 43
Expertise classification	14 / 22

3.2 Finger-tapping Experiment Results

As previously mentioned, 22 persons from the total polled answered to the survey (table 1). Results showed that 65% of the subjects had formal musical training with more than five years of experience. Around 30% were at least one-year experienced in percussion and 13% claimed to had taken some other kind of percussion training (minor education or self taught skills). We also considered other kinds of percussion experience and 43.5% reported to be familiar with drum machines or the manipulation of digital drums. Moreover, 65% declared to pay attention to rhythmic aspects when listening to music. Considering the outcome of this responses, as previously mentioned, from the 22 respondents we grouped 12 to be the non-experts in percussion (NEP), regardless of their music experience in other kinds of instrument, and 10 to the group of experts in percussion (EP).

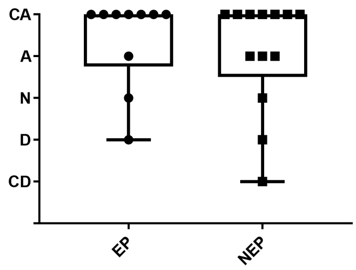
First, we manually estimated the tempo of the recordings using a regular DAW. Table 2 shows the mean and variance values of these estimations for each of the defined groups. Results revealed similar values in the average and variance BPM for both subgroups, ranging from 95 to 125 bpm, reinforcing previous research on human preferred tempo [10].

Table 2: Tempo estimation

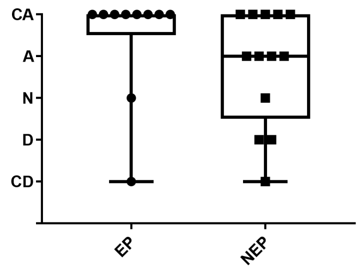
	NEP	EP
Average	109.5	110.0
Std. dev	14.3	15.0
Median	108.0	109.5

The analysis of the answers to the survey show some differences between EP and NEP. For each question in the survey, a Mann-Whitney U-test has been computed to assess the statistical significance of the observed differences between EP and NEP subjects. Because of the small amount of answers gathered, some interesting differences are just marginally significant and more data should be gathered in order to make strong claims about the observed trends. In figures 2a and 2b we can see minor differences between EP and NEP regarding their fingers and hands' usage for finger-tapping. A slight, but not statistically significant difference was observed in the sense that NEP used mainly their hands or, in some cases, their fingers, but rarely combined. Figure 2c shows the number of different percussion layers that were present in the pattern. NEP subjects, in general, were not able to play more than two. EP subjects played up to three different instrument layers in the pattern (p=0.171). As depicted in Figure 2d, EP subjects used the timbre characteristics of the object being tapped to provide differences to each stroke. Conversely, NEP users did not take into account this property of the surface (p=0.052). We can see in Figure 2e that most of EP subjects claimed that there is a match between groups of similar timbre strokes and a particular instrument layer, as opposed to NEP respondents, those who declared neutral or in disagreement with this statement. (p=0.014). Further results show that there was a trend among EP subjects to associate strokes played with particular parts of the hand with a certain type of percussion instrument (Figure 2g) (p=0.071). This fact, suggests that EP individuals took advantage of fingers' ergonomics to perform fast and successive strokes as in a snare-roll. Figure 2f reveals that neither EP nor NEP subjects tended to overlap strokes when finger-tapping (p= 0.667). This is an interesting result considering that many percussion instruments, such as a drum set, allow overlapping strokes.

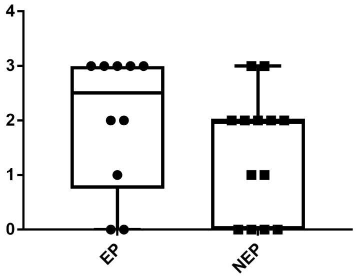
In conclusion, the results so far seem to suggest the existence of two general strategies for human finger-tapping based on the expertise: the NEP and EP profiles. The former subgroup is characterized by mainly using the hands for tapping, which constraints to two the number of percussion layers that they are able to play in the pattern. Also, NEP are not so concerned about timbre aspects when it comes to hit on different spots of the given surface so to confer timbre



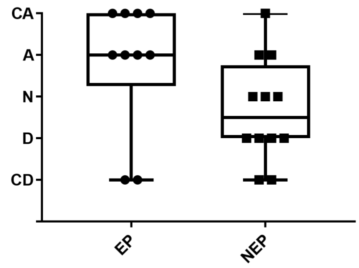
(a) "I used mainly my hands for tapping"



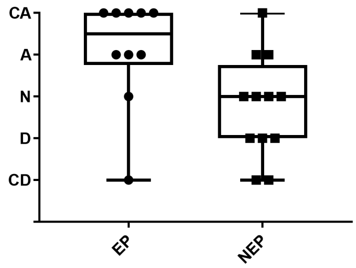
(b) "I used mainly my fingers for tapping"



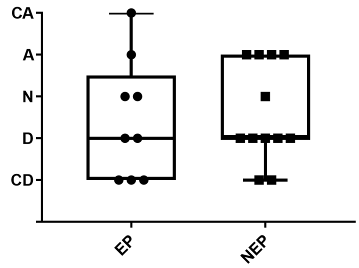
(c) "How many voices are present in the pattern?"



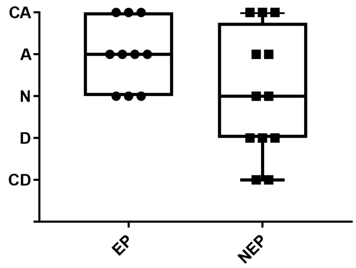
(d) "I made use of the timbre characteristics of the box to provide differences to each stroke"



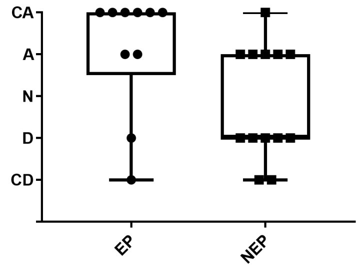
(e) "There is a match between each type of stroke and the voice it represents"



(f) "There are overlapping strokes in the pattern"



(g) "I used my fingers to perform as snares in a drum-set"



(h) "I was thinking on a percussion instrument to devise the pattern"

Fig. 2: Summary of responses to the questions (a-h)

differences to each stroke. Moreover, they do not think of each different stroke as a different voice but actually do play different percussion layers. In general, NEP do not think of a drum or a percussion instrument for conceiving the pattern. The EP profile is characterized by equally using hands and fingers for finger-tapping. In spite of using a more elaborated technique, members of this subgroup do not play more than three instrument layers in the pattern. Moreover, they take advantage of timbre to provide nuances to each tap as well as to associate these differences with a particular percussion layer. Even though they did not think on a drum to conceive the rhythm, they may imagine themselves as playing a general percussion instrument such as a "yembé" or a "cajón flamenco". In the following section we conducted a finger-tapping onset classification task in terms of the expertise level of the performer. The rationale for this experiment is that, if there are clear differences in the temporal and timbral features of both types of players, an automatic classifier based on those features should be able to tell apart onsets according to the level of expertise

3.3 Finger-tapping Expertise Classification

A dataset of 485 finger-tapping onsets was extracted from the 14 excerpts used for this experiments (table 1). We counted 327 onsets to be EP and 158 NEP. Onset detection was computed using Convolutional Neural Networks (CNN), following the method implemented in Madmom[7] python library [11]. For each of them, a set of 86 low-level descriptors, including those used in previous studies on percussive sound classification [12,13,14], were extracted using Essentia[8] python library. We used Weka[9] for running our experiments as well as the python library Scikit-learn[10], using 10-fold cross validation as evaluation procedure. We compared several machine learning algorithms and feature selection techniques, as well as attribute evaluators and search methods. In all the classification strategies that were explored, we obtained very high accuracies. Using Information Gain filter with a Ranker search yielded best hits, with an overall maximum of 96% using the Multi-layer Perceptron (MLP) algorithm (table 3).

Features ranked by the information gain filter belong to the spectral domain as shown in Figure 3. These descriptors are closely linked with the timbre characteristics of the sound, which suggests that it is on this music aspect where the main differences between both finger-tapping methods underlie.

4 Discussion

Results suggest the existence of two overall finger-tapping strategies: one addressed by people with no percussion training and another by people with at

[7] https://github.com/CPJKU/madmom
[8] http://essentia.upf.edu
[9] http://www.cs.waikato.ac.nz/ml/weka/
[10] http://scikit-learn.org/

```
Ranked attributes:
 0.436     72 sccoeff_5
 0.321     36 mfcc_5
 0.166     59 spectral_rolloff
 0.149     77 scvalley_4
 0.145     44 mfcc_13
 0.144     76 scvalley_3
 0.143      1 barkband_1
 0.14      78 scvalley_5
 0.131     41 mfcc_10
 0.128     71 sccoeff_4
```

Fig. 3: Attribute selection ranking with 10 best features computed in Weka

Table 3: Finger-tapping expertise level classification

K-best features	All	k=70	k=50	k=20	k=5
Naive Bayes	82.1	81.6	83.5	86.0	**86.4**
SVM	**95.9**	93.6	90.9	88.0	88.4
Logistic	**94.6**	91.1	89.1	88.6	86.6
Decision Tree	91.3	91.3	**92.0**	88.9	88.9
MLP	**96.1**	94.6	93.6	91.9	88.9

least two years' experience in some percussion activity. By analyzing the gathered data and the recorded audio content, we have described the behavior and characteristics of finger-tapping for each of them. In the case of NEP users, this finger-tapping strategy is characterized by not considering timbre to confer differences to each stroke when finger-tapping; they distinguish up to 2 different percussion layers in the pattern and they do not perform overlapping strokes. In the case of EP users, it has been observed that this group provides timbre nuances to each stroke, plays up to 3 percussion layers in the pattern and, in general, they finger-tap as if they were playing a percussion instrument. Unfortunately, not all the relevant differences between EP and NEP subjects reached statistical significance. Even considering that, these differences suggest that a potential software for transcribing finger-tapping could have different implementation approaches depending on the music expertise of the user.

5 Conclusions and Future Work

We have presented an exploratory study on the behavioral and acoustic properties of human finger-tapping. This work has been drawn towards the implementation of new audio-driven interfaces for drum composition, capable of using this human behavior as a means for interaction. To this end we have accomplished different sub-goals. First, we have collected a data-set of 43 finger-tapping recordings performed by individuals with different music backgrounds. Secondly, we have analyzed the on-line answers given by 22 subjects about their finger-tapping performance. In this survey we also collected user-context information by requesting people's preferences and needs with regard to a hypothetical interactive software for the arrangement of drum-sets. Thirdly, we detected onsets and computed several low-level descriptors from the recordings. Analysis over the survey results and the extracted audio features suggested the existence of two general strategies for finger-tapping: one followed by people with no experience at all in percussion (NEP) and another one typical of people with at least two years of experience in some percussion facet (EP). Automatic classification experiments allowed us to get support for these two strategies, based on different (mostly) temporal and spectral descriptors of the hits generated by the participants in the finger-tapping task. Our classifier was able to tell apart EP and NEP strokes with an accuracy of 96%.

Directions for future work include to increase the training examples for the classification task. Also, to approach experiments from a functional perspective (i.e. for the purpose of implementing an audio-driven drum arrangement software) rather than in a general and spontaneous perspective, as the followed in this study. For example, we propose to submit a group of volunteers to listen a certain drum pattern and, after some time, ask them to perform it by finger-tapping. To our knowledge, this method would provide more accurate results in the context of finger-tapping transcription.

Acknowledgments. This research has been partially supported by the EU funded GiantSteps project (FP7-ICT-2013-10 Grant agreement nr 610591).

References

1. G. Peters, C. Anthony, and M. Schwartz, "Song search and retrieval by tapping," *Proceedings of the National Conference on Artificial Intelligence*, p. 1696, 2005.
2. P. Hanna and M. Robine, "Query by tapping system based on alignment algorithm," *ICASSP '09 Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1881–1884, 2009.
3. C. Chen and J. Jang, "A Shifted Alignment Algorithm For Query By Tapping," *MIREX 2013*, pp. 701–705, 2013.
4. S. Granqvist and S. Dahl, "Looking at perception of continous tempo drift - A new method for estimating internal drift and just noticeable difference." *Proceedings of the Stockholm Music Acoustics Conference*, vol. 2003, no. Smac 03, pp. 595–598, 2003.
5. W. T. Fitch and A. J. Rosenfeld, "Perception and Production of Syncopated Rhythms," pp. 43–58, 2007.
6. S. Dixon and W. Goebl, "Pinpointing the beat: Tapping to expressive performances," *To appear in: 7th International Conference on Music . . .* , no. July, pp. 2000–2003, 2002.
7. M. C. Tiffin-Richards, M. Hasselhorn, M. L. Richards, T. Banaschewski, and A. Rothenberger, "Time reproduction in finger tapping tasks by children with attention-deficit hyperactivity disorder and/or dyslexia," *Dyslexia*, vol. 10, no. 4, pp. 299–315, 2004.

8. P. Arias, V. Robles-García, N. Espinosa, Y. Corral, and J. Cudeiro, "Validity of the finger tapping test in Parkinson's disease, elderly and young healthy subjects: Is there a role for central fatigue?" *Clinical Neurophysiology*, vol. 123, no. 10, pp. 2034–2041, 2012.
9. M. Franek, J. Mates, T. Radil, and E. Beck, K.; Pöppel, "Finger tapping in musicians and nonmusicians," *International Journal of Psychophysiology*, vol. 11, no. 3, pp. 277–279, 1991.
10. D. Moelants, "Preferred tempo reconsidered," *Proceedings of the 7th international conference on music perception and cognition*, pp. 580–583, 2002.
11. J. Schlüter and B. Sebastian, "Musical Onset Detection with Convolutional Neural Networks," *6th International Workshop on Machine Learning and Music (MML)*, pp. 1–4, 2014.
12. P. Herrera, A. Yeterian, and F. Gouyon, "Automatic classification of drum sounds: a comparison of feature selection methods and classification techniques," *Music and Artificial Intelligence*, pp. 69–80, 2002.
13. P. Herrera, A. Dehamel, and F. Gouyon, "Automatic Labelling of Unpitched Percussion Sounds," *Aes 114Th Convention*, 2003.
14. P. Herrera, V. Sandvold, and F. Gouyon, "Percussion-related semantic descriptors of music audio files," *Proceedings of 25th International AES*, pp. 1–5, 2004.