# Towards a comprehensive dataset of vocal imitations of drum sounds

MEHRABI, A; Dixon, S; Sandler, M; 2nd AES Workshop on Intelligent Music Production

# TOWARDS A COMPREHENSIVE DATASET OF VOCAL IMITATIONS OF DRUM SOUNDS

*Adib Mehrabi, Simon Dixon and Mark Sandler*

Centre for Digital Music
Queen Mary University of London
{a.mehrabi,s.e.dixon,mark.sandler}@qmul.ac.uk

## ABSTRACT

The voice is a rich and powerful means of expressing acoustic concepts such as musical sounds. Recent research on vocal imitations has demonstrated the viability of using the voice to search for sounds, using query by vocalisation. Here we present the methods used to develop a dataset for evaluating the performance of query by vocalisation systems for drum sounds. The dataset consists of imitations of 30 drum samples from a commercial drum sample library, performed by 14 musicians with experience in computer based music production. The dataset includes participant ratings of their satisfaction with each imitation, and perceptual similarity ratings between each imitation and the sounds being imitated, collected via an online, MUSHRA style listening test.

## 1. REQUIREMENTS OF THE DATASET

Searching for drum sounds is a core part of the electronic music making process [1], and the voice is an effective means of describing sounds [2]. Query by vocalisation (QBV) systems allow a user to search for a sound by vocalising an example of the desired sound [3,4]. This interaction modality presents an intuitive way for musicians, music producers and sound engineers to search for musical sounds using intelligent search methods. However, to build a QBV system that can retrieve perceptually relevant sounds, we require a model that maps between the sound spaces of the voice and a sample library, based on *a priori* knowledge of the perceptual similarity between vocal imitations and the samples in question. To design and build such a model, we require a dataset of prototypical vocal imitations that includes perceptual similarity ratings between the imitations and each of the sounds being imitated. The primary aim of this work is to develop such a dataset, specifically for drum samples, however the methods used here could also be applied to other types of sample libraries.

## 2. SELECTING THE STIMULI

The drum samples were selected from the *fxpansion*[1] *BFD3 Core* and *8BitKit* sample libraries, with six samples taken from each of five drum classes (kicks, snares, hi-hats, toms, cymbals), giving thirty samples to be used as the stimuli. The samples for each class were selected as follows: first, a random seed sample was selected; next, the most and least similar samples to the seed were selected, based on a within-drum-class similarity measure using auditory images [5]; finally, three samples equally spaced in distance between the closest and furthest samples were selected. This approach gives a range of six samples within a drum class that are representative of the variety of sounds in the sample library.

The auditory image based similarity measure is an implementation of the best performing method in [5], which the authors found to be highly correlated with perceptual similarity ratings of within-class drum sounds for bass, snare and tom drum classes. In brief, this measures the distance between the spectrograms of two drum sounds after after the following pre-processing: length is matched by zero padding the shorter spectrogram; loudness (in dB) is scaled using Terhardt's ear model [6]; frequency scaled using the Bark scale.

## 3. RECORDING THE IMITATIONS

Fourteen participants recorded their vocal imitations of the thirty extracted samples. The recording workflow is shown in Figure 1. The participants could practise and re-record the imitations of each stimulus as many times as they wished. After recording each imitation, the participants gave satisfaction ratings for their imitations on a five point Likert scale from completely dissatisfied to completely satisfied.

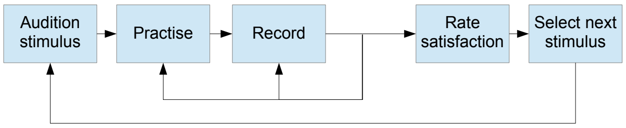


Figure 1: Workflow for recording each vocal imitation.

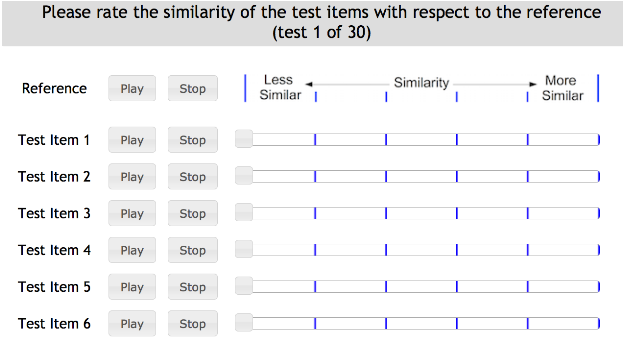
## 4. LISTENING TEST FOR SIMILARITY RATINGS

In the second part of this work, we conducted an online MUSHRA[2] style listening test to collect similarity ratings for each imitation with respect to the stimuli from the same class. The implementation of this test is based on BeaqleJS [7]. Each participant is presented with 28 test pages taken from a random selection of the 420 imitations, plus 2 repeated imitations. The purpose of the repeated imitations is to ensure participants are able to repeat their ratings when presented with the same task twice. For each test page, the participant is presented with an imitation as the reference

[1] https://www.fxpansion.com

[2] https://www.itu.int/rec/R-REC-BS.1534/en

sound and 6 test items, which are the 6 drum samples from the class of the imitated sample, including the imitated sample as the hidden reference. For each imitation, the similarity between the imitation and each of the within-class drum samples is measured on a relative scale from ‘less similar’ to ‘more similar’, as shown in Figure 2.



Please rate the similarity of the test items with respect to the reference (test 1 of 30)

Figure 2: Example of a single test page from the listening test.

## 5. CONCLUSION

We have created a dataset of vocal imitations of 30 drum samples, which were taken as a representative sample of kicks, snares, hi-hats, toms and cymbals from a commercial sample library. To complement this dataset we have conducted an online listening study to measure the similarity between each imitation and the samples from the drum class of the imitated sound. This provides a comprehensive dataset that can be used to evaluate different QBV models. At the time of writing, data is still being collected for the listening test and we encourage readers to partake in this part of the study[3].

## 6. REFERENCES

[1] K. Andersen and F. Grote, “Giantsteps: Semi-structured conversations with musicians,” in *Proc. 33rd Annu. ACM Conf. Extended Abstracts on Human Factors in Computing Systems*, Seoul, 2015, pp. 2295–2300.

[2] G. Lemaitre and D. Rocchesso, “On the effectiveness of vocal imitations and verbal descriptions of sounds,” *J. Acoustical Soc. America.*, vol. 135, no. 2, pp. 862–873, 2014.

[3] Y. Zhang and Z. Duan, “Retrieving sounds by vocal imitation recognition,” in *Proc. 25th IEEE Workshop on Machine Learning for Signal Processing*, Boston, 2015, pp. 1–6.

[4] D. S. Blancas and J. Janer, “Sound retrieval from voice imitation queries in collaborative databases,” in *Proc. 53rd AES Conf. on Semantic Audio*, London, 2014, pp. 2–8.

[5] E. Pampalk et al., “Hierarchical organisation and visualisation of drum sample libraries,” in *Proc. Digital Audio Effects Conf. (DAFx-04)*, Naples, 2004, pp. 378–383.

[6] E. Terhardt, “Calculating virtual pitch,” *Hearing Research.* vol. 1, no. 2, pp. 155–182.

[7] S. Kraft and U. Zölzer, “Beaqlejs: Html5 and javascript based framework for the subjective evaluation of audio quality,” presented at Linux Audio Conf., Karlsruhe, 2014.

[3] http://am-drumstudy.apps.devcloud.eecs.qmul.ac.uk/