

## Perceptual Evaluation of Synthesised Sound Effects

David Moffat<sup>1\*</sup> and Joshua D. Reiss<sup>1</sup>

<sup>1\*</sup>Center for Digital Music, Queen Mary University of London, UK, [d.j.moffat@qmul.ac.uk](mailto:d.j.moffat@qmul.ac.uk)

**Abstract**— Sound synthesis is the method of artificially producing sounds. A range of sound effects were generated through the use of five different synthesis techniques, across eight sound classes to produce sixty six different example audio samples. Perceptual evaluation was performed to identify the perceived realism of each of the synthesis techniques. The results show that some synthesis techniques can be considered as realistic as a recorded sample, when listening directly to audio samples.

### I. INTRODUCTION

Sound Synthesis (SS) is the technique of generating sound through artificial means. Synthesis has been demonstrated in the area of sound effect, which can be used in production of a range of popular media, such as video games, TV, film and augmented or virtual reality. A listening test was undertaken in which participants were asked to rank audio files in terms of their perceived realism.

### II. METHOD

Synthesis of a range of sound effects was performed through a range of methods including statistical modelling, sinusoidal modelling, sinusoidal modelling with residual, physically informed modelling, granular synthesis, concatenative synthesis and additive synthesis.

Thirteen participants were asked to evaluate sounds for eight categories (applause; babble; bees; fire; rain; stream; waves; wind) and in every category between six and eleven samples were provided. Sixty six samples were evaluated in total. All samples were loudness normalised. Each category had at least some anchor and at least one recorded sample. The recorded samples were all selected by a group of five experienced critical listeners as being realistic samples. Participants were asked to rate all samples within a given category, relative to all the other samples within the category. They were to rate the samples relative to how realistic they perceived the sounds on a continuous linear scale labelled from "very unrealistic" through "quite unrealistic", "quite realistic" to "very realistic". The provided synthesised samples were also compared to at least one reference and at least one anchor per category, similar to a mushra style test. Participants did not have any information regarding the samples, other than a reference number, which was randomised. The initial locations of samples were randomised and the order of categories participants were evaluating was randomised. The listening test was constructed using the Web Audio Evaluation Toolbox [1].

### III. RESULTS

ANOVA was undertaken to determine if the distributions of the results are the same. SPAD [2] was the only synthesis method that cannot be considered statistically significant from the reference audio samples. It can be considered as realistic as a recorded sample in the case of all sound classes. The user ratings are presented in Figure 1. The red line represents the median. The end of the notches represents the 95% confidence intervals, and the end of the boxes represent the 1st and 3rd quartiles. The end of the whiskers represent the range of the data not considered as an outlier. Red crosses are outliers.

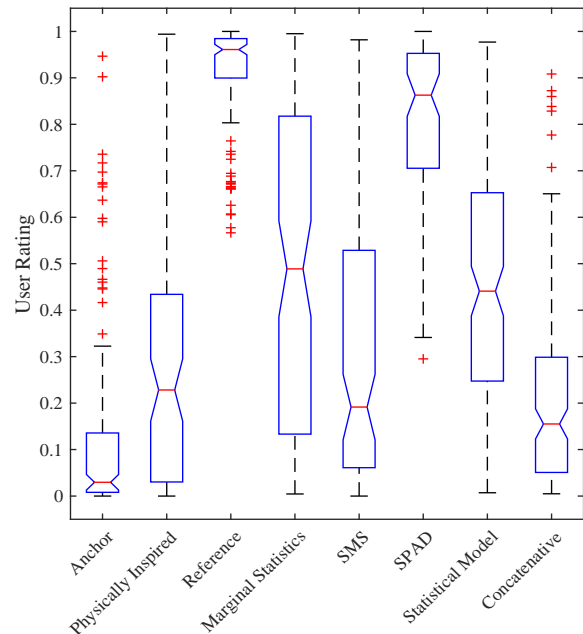


Figure 1. Perceived Realism of all Synthesis Methods

### REFERENCES

- [1] N Jillings, B De Man, D Moffat and JD Reiss, "Web Audio Evaluation Tool: A browser-based listening test environment" in Proc. International Sound and Music Computing Conference (2015).
- [2] Verron, C., Aramaki, M., Kronland-Martinet, R., & Pallone, G. (2010). A 3-D immersive synthesizer for environmental sounds. IEEE Transactions on Audio, Speech, and Language Processing, 18(6), 1550-1561.