*Proceedings of the 3rd Workshop on Intelligent Music Production, Salford, UK, 15 September 2017*

# AUDIO PROCESSING CHAIN RECOMMENDATION USING SEMANTIC CUES

*Spyridon Stasis, Nicholas Jillings, Sean Enderby and Ryan Stables*

Digital Media Technology Lab
Birmingham City University
{spyridon.stasis,nicholas.jillings,sean.enderby,
ryan.stables}@bcu.ac.uk

## ABSTRACT

Sound engineers typically allocate audio effects to a channel strip in series. This allows the engineer to perform a complex set of operations to fine-tune different tracks in a mixing or mastering environment. In this research, trends in plugin chain selection are investigated, focusing on transformations which modify the timbral characteristics of a sound. Using this information, a recommendation system can be constructed to generate full processing chains in a Digital Audio Workstation (DAW).

## 1. INTRODUCTION

### 1.1. Background

Recording engineers perform a series of complex processing tasks, either for creative or corrective reasons. One of the more demanding tasks, requiring a relatively high level of experience, is the selection and ordering of effects in a processing chain. This process involves the combined configuration of various effects in series for each channel strip in the mix. If nonlinear effects such as compression and distortion are used, the order of effects becomes very important due to the non-commutable nature of these systems. In this study, we investigate the application of processing chains to a predefined set of music mixes, with respect to a corpus of descriptive terms. The resulting system aims to bridge the gap between experienced sound engineers and novice users [1].

### 1.2. Previous Work

The use of processing chains in Intelligent Music Production (IMP) has been investigated previously in the context of automatic mixing. Pestanta [2] for example discusses the difficulty of comparing plugin sequences, in which interview findings regarding the placement of audio effects in a specific order are presented. Similarly, in an attempt to bypass the problem of effect placement, Wise [3] proposes a single processing tool that combines the manipulation of frequency components and dynamics using a single effect. In this study we build on the work conducted in [4], in order to utilise descriptive terminology for processing chain recommendation. This allows us to use the perceived timbral effects of a processing chain as a method of comparative evaluation.

## 2. METHODOLOGY

To analyse of the production decisions that a sound engineer makes when selecting effects in a processing chain, an experiment was conducted where participants were asked to use a series of $N \geq 1$, cascaded audio effects to process a set of musical stimuli according to a predefined semantic descriptor. In total 178 submissions were made by 47 individuals, all of whom had experience in music production. The subjects were provided with four audio effects (EQ, compressor, distortion and reverb), with no restrictions on the length of their processing chain, the number of instances of a particular effect in the chain, or the ordering of the effects. The test was performed online, in a browser-based DAW [5], where all the effects were built using the JSAP web audio plugin framework [6].

The stimuli were selected from the *Mixing Secrets* library[1], obtained via the Open Multitrack Testbed [7], to feature an array of different instruments across five genres. The instruments selected were *acoustic guitar*, *bass guitar*, *drums (mixed)*, *electric guitar*, *piano*, *saxophone*, *violin*, and *vocals*. To further evaluate the processing chain choice in a mastering scenario, *complete mixes* were also used. All of the audio samples presented to the subjects had a duration of 30 seconds, and in the case of the *complete mixes*, all the instruments were active at the excerpt selected.

The timbral adjectives given to participants were derived from analysis of the SAFE Project [8, 9]. Here, the terms were selected for their generalisability across effects. These were *air*, *boom*, *bright*, *close*, *cream*, *crisp*, *crunch*, *damp*, *deep*, *dream*, *fuzz*, *punch*, *room*, *sharp*, *smooth*, *thick*, *thin*, and *warm*.

## 3. RESULTS

In total, 30 unique processing chains were collected during the experiment, and 11 of these were excluded due to them appearing only once in the dataset. From the remaining 19 chains (including single effects), the most popular configurations are EQ (27.5%), reverb (12.5%), compressor-EQ

[1] Available at http://www.cambridge-mt.com/ms-mtk.htm

(11.9%), distortion (8.9%), EQ-compressor (8.9%) and EQ-reverb (5.3%). Given the results, it is possible to represent the relationship between the descriptors by treating each processing chain as a separate dimension. In this context the entries of the high-dimensional dataset are the timbral adjectives, and the dimensions correspond to the processing chains accumulated through the data gathering phase. The dimensions are weighted with regards to the chain's use for achieving a term, and are in-turn normalised. Principal Component Analysis (PCA) can then be performed to project the datapoints into a low-dimensional space. Following the process of dimensionality reduction, hierarchical clustering is implemented to organize the resulting structure of the PCA mapping into groups. Figure 1 highlights the existence of three predominant groups: one that uses mainly reverb (*room*, *damp*, *dream*), one that uses mainly distortion (*fuzz*, *punch*, *crunch*), and a larger group for terms that can be achieved through a wider range of plugin chains.

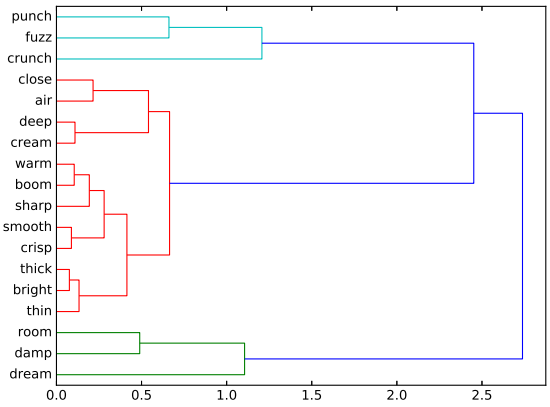


Figure 1: *Hierarchical clustering of unique terms based on processing chain usage.*

## 4. PROCESSING CHAIN RECOMMENDATION

The processing chain recommendation system can be implemented using a Markov Model [4, 10, 11]. This allows for the probabilistic selection of a plugin from a finite number of states, given the previous state in the chain. As the model is trained using the processing chain data collected during our experiment, these conditional probabilities are intended to represent the selections of expert users. To make the model specific to a given timbral adjective, a state transition matrix based on the probabilities of each descriptive term is used. For example, *dream* will generate a chain of only reverb with a likelihood of 49.38% or one consisting of EQ-reverb with a likelihood of 29.63%.

Using the chains generated by the model, we can compare the similarity of terms in our dataset. This is illustrated in Figure 2, in which n-dimensional chains are projected into a 2-dimensional space using PCA. Here, descriptors that can be achieved through similar plugin chains are placed close together, as is the case with *smooth*, *thick* and *boom*, or *room* and *damp*.

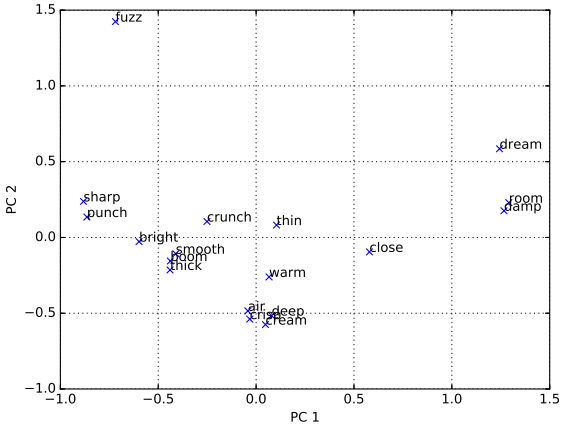


Figure 2: Low-Dimensional Mapping for the Markov Chain recommender

## 5. OBJECTIVE EVALUATION

To evaluate the performance of the recommender system, we attempt to measure its performance at retaining the structure of the original space. This is a well-known problem in dimensionality reduction, where it is necessary to retain the structure of the high-dimensional space in the low-dimensional mapping [12, 13]. The similarity of the original and generated descriptor mappings can be evaluated through the *trustworthiness* ($T_k$) and *continuity* ($C_k$) metrics [12], shown in Eqs. 1 and 2. The metrics perform a rank-based comparison between the two spaces for a varying number of neighbours. The distances of the $n$ entries in two spaces ($U$ and $V$) are converted to ranks ($r$ and $\hat{r}$) between points $i$ and $j$. The measurements then evaluate the distributions of datapoints in the respective spaces over a number of neighbouring datapoints ($k$).

$$T_k = 1 - \frac{2}{nk(2n - 3k - 1)} \sum_{i=1}^{n} \sum_{j \in U_i^{(k)}} (r(i,j) - k) \quad (1)$$

$$C_k = 1 - \frac{2}{nk(2n - 3k - 1)} \sum_{i=1}^{n} \sum_{j \in V_i^{(k)}} (\hat{r}(i,j) - k) \quad (2)$$

The Markov chain space achieves a *trustworthiness* score of 0.78 for the original structure of unique terms, suggesting that the organisation of descriptors is retained, and a score of 0.782 for *continuity*.

## 6. CONCLUSION

We present an analysis of the trends in audio processing chains, based on perceived timbral transformations. To do this, we construct a novel processing chain recommender using a Markov Chain model. We measure the performance of the system through structure preservation metrics and show that it is capable of preserving the original relationships between descriptors with a reasonable degree of accuracy.

## 7. ACKNOWLEDGMENTS

The work of the first author is supported by The Alexander S. Onassis Public Benefit Foundation.

## 8. REFERENCES

[1] J. Reiss, "Intelligent systems for mixing multichannel audio," in *17th International Conference on Digital Signal Processing*, pp. 1–6, July 2011.

[2] P. D. L. G. Pestana, "Automatic mixing systems using adaptive digital audio effects," 2013.

[3] D. K. Wise, "Concept, design, and implementation of a general dynamic parametric equalizer," *Journal of the Audio Engineering Society*, vol. 57, no. 1/2, pp. 16–28, 2009.

[4] S. Stasis, N. Jillings, S. Enderby, and R. Stables, "Audio processing chain recommendation," in *Proceedings of the 20th International Conference on Digital Audio Effects*, (Edinburgh, UK), 2017.

[5] N. Jillings and R. Stables, "Investigating music production using a semantically powered digital audio workstation in the browser," in *Audio Engineering Society Conference on Semantic Audio*, (Erlangen, Germany), June 2017.

[6] N. Jillings, Y. Wang, J. D. Reiss, and R. Stables, "JSAP: A plugin standard for the Web Audio API with intelligent functionality," in *Audio Engineering Society Convention 141*, 2016.

[7] B. De Man, J. D. Reiss, *et al.*, "The open multitrack testbed: Features, content and use cases," in *Proceedings of the 2nd AES Workshop on Intelligent Music Production*, vol. 13, 2016.

[8] R. Stables, S. Enderby, B. De Man, G. Fazekas, and J. D. Reiss, "SAFE: A system for the extraction and retrieval of semantic audio descriptors," in *15th International Society for Music Information Retrieval Conference*, 2014.

[9] R. Stables, B. De Man, S. Enderby, J. D. Reiss, G. Fazekas, and T. Wilmering, "Semantic description of timbral transformations in music production," in *ACM on Multimedia*, pp. 337–341, 2016.

[10] A. Markov, "Extension of the limit theorems of probability theory to a sum of variables connected in a chain," *Dynamic Probabilistic Systems*, vol. 1, 1971.

[11] G. Tauchen, "Finite state markov-chain approximations to univariate and vector autoregressions," *Economics Letters*, vol. 20, no. 2, pp. 177 – 181, 1986.

[12] S. Kaski, J. Nikkilä, M. Oja, J. Venna, P. Törönen, and E. Castrén, "Trustworthiness and metrics in visualizing similarity of gene expression," *BMC bioinformatics*, vol. 4, no. 1, p. 48, 2003.

[13] L. Van Der Maaten, E. Postma, and J. Van den Herik, "Dimensionality reduction: a comparative," *J Mach Learn Res*, vol. 10, pp. 66–71, 2009.