Assisted Parameter Modulation in Music Production using Large-Scale Producer-Defined Semantics

Ryan Stables[1] , Matthew Cheshire[2]

[1]Anglia Ruskin University, East Road, Cambridge, UK, CB1 1PT.
ryan.stables@anglia.ac.uk
[2]DMT Lab, Birmingham City University, Birmingham, UK, B47XG.
matthew.cheshire@mail.bcu.ac.uk

1. Abstract

In music production, audio effects and synthesis parameters often address technical low-level processes, which can be difficult to interpret and use without experience and extensive training. In order to make the process more accessible, we therefore investigate methods of creating high-level parameters, thus bridging the gap between musicians and technology. In this study we develop semantic parameters for audio effects by gathering large amounts of musical semantics data from music-producers at the point of content creation. By extracting the parameter state of an audio plug-in and a series of audio features, we can then use the data to non-linearly modulate parameters within the same interface.

2. Introduction

In music production, audio effects and synthesizers are controlled using parameters that interface with statistical, low-level properties of the audio signal. In order for users to achieve desired musical results, they need to be familiar with this low-level parameter space. This can require extensive training, which typically results in a separate engineer being employed for music production tasks. Generally, the desired characteristics of sounds can be described using semantic or high-level descriptors. For example, if a guitar has a significant amount of high-frequency energy, it is often described as having a bright timbre. These perceptual descriptors are not always defined statistically, and often have some non-linear relationship to the low-level parameter space.

In order to improve access to music production, it is useful to interpret the relationship between these low and high-level parameter spaces. This allows us to design audio effects and synthesizers that can be controlled with semantically understandable parameter spaces, thus providing musicians with a way to intuitively manipulate audio. In this study we present a methodology for the capture of representative high and low-level descriptors during the music production workflow, followed by a model for the mapping of large-scale semantics data to local audio-unit parameter spaces.

The relationship between low-level audio features and semantic descriptors has previously been estimated using laboratory-based subjective listening tests. [1, 2] for example, consider the relationship between audio features and perceived timbre spaces using ranked samples, taken from around 50 subjects. In [1] correlation and hierarchical-cluster analysis are used to identify feature salience, whereas [2] uses metric and non-metric data reduction techniques to identify salient dimensions in the data. Whilst these experiments yield useful results, the tests are undertaken in laboratory conditions, with forced predefined word selections. In an alternative approach to the laboratory-based exercises, [3] uses a web-based gaming interface to collect musical metadata, here the data is used to test a machine learning algorithm. This study allows for more representative data, however the tags are predetermined and there is little motivation for the user-base to utilize the system.

3. Methodology

To ensure the data is both representative of the music production community, large amounts of labeled data is taken from within the Digital Audio Workstation (DAW), thus leading to a model for the estimation of perceptually accurate descriptors based on a large corpus of semantically annotated music production data. To compile the dataset, we have developed a series of DAW plug-ins, designed to non-linearly map semantics data to low-level parameters. These devices encourage users to continually label parameter-states at the point of content creation, leading to transferable semantic profiles, each of which can be applied independently of the production tool they were recorded in. The plug-ins will be developed to automatically return the anonymous data, thus continually updating the dataset with parameter-states, acoustic features and semantic labels.

This methodology is intended to gather an extremely large corpus of data, to mitigate against bias and to provide a more representative account of personalized semantic descriptors. The data then allows users to quantify their own perception of musical signals and use the self-adapting tools provided to map this to the production process. To analyze the data, we use techniques taken from the fields of Semantic Web, Natural Language Processing and Machine Learning to partition the user-data into groups with acoustic, linguistic and musical similarity, providing a structure, which allows relationships to be inferred within the grouped feature space.

To partition the dataset further, our system allows for the population of meta-data fields from within the plug-in interface. This means we can capture additional information from the user such as location, age, sex, genre, and instrumentation. These attributes are then used to identify feature subsets, allowing us to evaluate the variance of semantic descriptors over different conditions.

4. Architecture

The plug-ins are developed to allow users to both load and save parameter states from within the same UI. This encourages users to utilise the plug-ins for production purposes, but also exercises the importance of saving states. Once a semantic descriptor has been saved, the text is sent to the server along with a set of features, extracted from the audio and the values from the parameter space. The feature set used for audio analysis is defined by the LibXtract library [4], this is an audio feature extraction framework, which includes around 40 different features, taken from 10 different input representations. These include temporal features such as Log Attack Time, Spectral features such as spectral centroid and Kurtosis and augmented feature vectors such as MFCCs.

The feature data is then processed on the server using Multidimensional Scaling to reduce the dimensionality of the feature set and the semantic terms are clustered using Natural Language processing. The objective is to identify correlation in similar terms, allowing parameters to be modulated. When the data is processed, it is stored in a separate database, allowing users to load the semantics information back into the plug-in using the same text field. This is shown schematically in Figure 1.

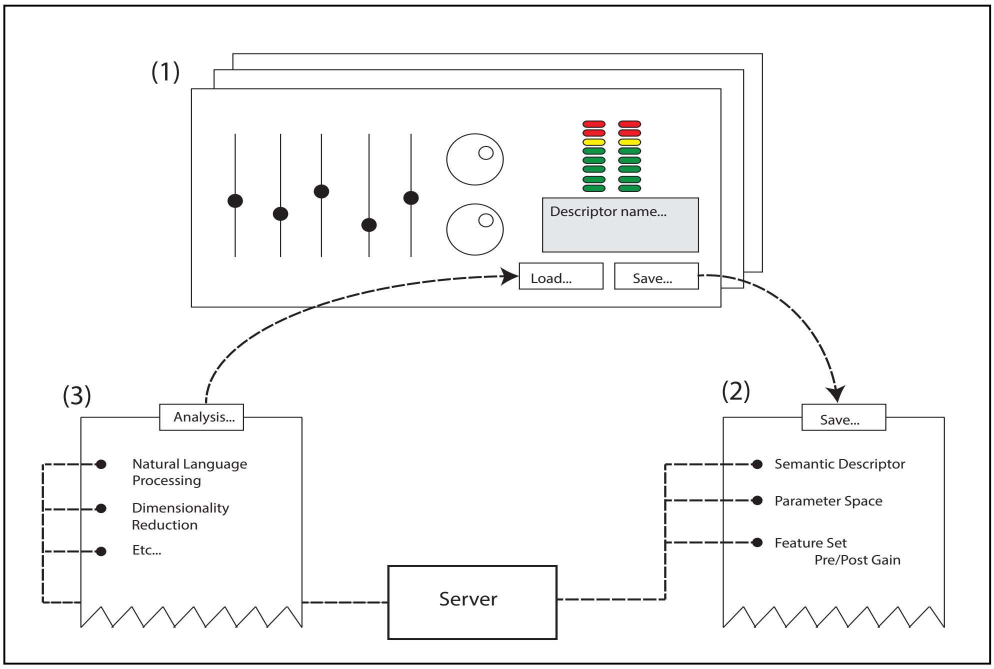

Figure 1: Schematic overview of the plug-in architecture

5. Conclusion

We have developed a system for the large-scale capture of producer-defined musical semantics by integrating the data-capture system into the DAW workflow. The system captures the parameter space, semantic descriptions of the sound and a set of audio features. These attributes are sent to a server, processed and stored in order to influence further production decisions. The data is processed using dimensionality reduction and NLP techniques, and the type of parameter modulation can be refined based on user-defined meta-data partitions.

6. References

[1] Disley, A., Howard, D., and Hunt. A. "Timbral description of musical instruments." *International Conference on Music Perception and Cognition*. 2006.

[2] Zacharakis, A., Pastiadis, K., Reiss, J. D., & Papadelis, G. Analysis of Musical Timbre Semantics through Metric and Non-Metric Data Reduction Techniques. *International Conference on Music Perception and Cognition*. 2012

[3] Mandel, M I., and Ellis, D. "A web-based game for collecting music metadata." *Journal of New Music Research* 37.2 (2008): 151-165.

[4] Bullock, J. "LibXtract: A lightweight library for audio feature extraction." *Proceedings of the International Computer Music Conference*. Vol. 43. 2007.