SPATIALLY RENDERING DECOMPOSED RECORDINGS – INTEGRATING SCORE-INFORMED SOURCE SEPARATION AND SEMANTIC PLAYBACK TECHNOLOGIES

Florian Thalmann, Sebastian Ewert, Mark Sandler, Geraint A. Wiggins
Queen Mary University of London

ABSTRACT

In this contribution, we present a system for creating novel renderings of a given music recording that aurally highlight certain musical aspects or semantics using spatial localizations. The system decomposes a monaural audio recording into separate events using score-informed source separation techniques and prepares them for an interactive mobile player that renders audio based on semantic information. We demonstrate the capabilities of the system by means of an example using an immersive chroma helix model which the listener can navigate in realtime using mobile sensor controls.

1. INTRODUCTION

Music has often been perceived as the most abstract of arts. In particular, while audio recordings capture musical and acoustical details of a piece, structural and representational information typically remain latent and thus difficult to access. For students learning about music and its structure, this can be particularly challenging as understanding also means identifying structural information in a recording and bringing it in line with the auditory perception. One way of dealing with this problem is visualization, which explains the emergence of numerous visualization tools in recent years, e.g. [1, 4]. However, there are several challenges in visualizing music. First, cross-modal mappings between auditive structures and their visual representation are often not intuitive and need to be learned as well, such that many different possibilities have been proposed in the past, each of them having individual advantages and disadvantages. Second, due to our visucentric nature, such visualizations may distract from the music or the listening process itself and thus what is being learnt might be different from an implicit expectation.

In this demonstration, we investigate the technical and educational potential of spatially sonifying music based on semantic content by combining two music technologies in a novel way. More specifically, we employ score-informed source separation techniques [2] to first decompose a given recording into its constituent note events. Using the structurally rich information provided by the score, we automatically annotate the note events with semantic classes. By associating these classes with parameters in our semantics-driven player framework, we can render each note event depending on its class and thus highlight the class aurally in a new sound mix. We demonstrate these capabilities using pitch-based classes associating each class with a different location along Shepard's pitch helix [1] (Figure 1) or similar models in a 3D space and render the resulting spatial mix binaurally. This way, we enrich the existing perceptual pitch information in the recording with spatial information, with the angle encoding the chroma and the height the octave of a note. Depending on the spatial distribution various musical concepts can be highlighted (or attenuated) in perceived geometrical patterns – intuitively, without explicitly learning an assignment as in visualizations. The demo will run on a smartphone or tablet with headphones, enabling the user to move within the helix and change its extent, adding control and interactivity to the focus on specific pitch classes appearing in the piece.

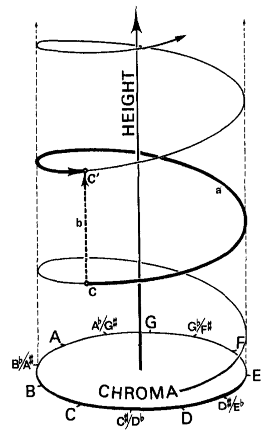

Figure 1. The pitch helix as illustrated in [3, p. 308].

2. SCORE-INFORMED SOURCE SEPARATION

With many advancements made in recent years, the decomposition of a musical recording into individual sound

[1] The pitch helix was first suggested by Drobisch in 1855 and later generalized by Shepard for other closely-related intervals, such as perfect fifths [3].

sources, a task often referred to as source separation, has become a central topic in music information retrieval and processing. Applications range from remixing tools, over karaoke generation, denoising and repair methods to automated upmixing and pre-processing in MIR. The task, however, is and remains highly challenging, and without additional prior knowledge it is mathematically ill-posed. Many different types of prior knowledge have been evaluated in recent years, ranging from assumptions about spectro-temporal properties of instruments to having the user manually select parts of an instrument in a spectrogram representation. A final solution, however, has yet to be found. A very promising direction is to use information provided by a musical score, which often leads to separation results of a quality high enough for various applications [2]. In particular, knowledge of which notes are being played enables a decomposition on a note rather than the usual instrument level, which allows for a more fine-grained separation. At the same time, each note can be annotated with pitch, duration or velocity information as available from the score, which can be used by our player framework. In this demonstration, we employ a method based on non-negative matrix factorization where the score information is translated into constraints for the underlying parameter estimation process, see [2] for more details. It employs a straightforward temporal evolution model, using different spectral representations for the onset and sustain parts of notes to increase the separation quality.

3. RENDERING USING THE SEMANTIC PLAYER

The Semantic Player is a cross-platform mobile application made with Ionic [2], ngCordova [3], JavaScript, the Web Audio API [4] as well as Semantic Web technologies, and it is designed to play back music in *indeterminate*, *context-dependent*, and *interactive* ways. It is based on a musical format called *Dynamic Music Objects (DMOs)*, an amalgamation of audio files, a structural definition including analytical and semantic information, and a rendering definition, which associates various mobile controls and autonomous control units with musical parameters. The player generates its interface and sensor allocations on the fly for each DMO configuration. Such DMOs can be prepared using the DMO Designer web application which allows users to load any kind of audio features and construct a DMO structure along with a rendering, based on these features. The output of the DMO Designer, an ontological definition in Semantic Web format, can directly be fed into the player along with the audio files and played back.

In the case of our experiment, we render the spatial position (x, y, z) of each note event as a function of its midi pitch p as follows:

$$(x, y, z) = (\cos(2\pi p'), \sin(2\pi p'), p/12)$$

with $p' = (p \mod 12)/12$. This is implemented using so-called feature mappings from DMO features to some

[2] http://ionicframework.com
[3] http://ngcordova.com
[4] http://www.w3.org/TR/webaudio/

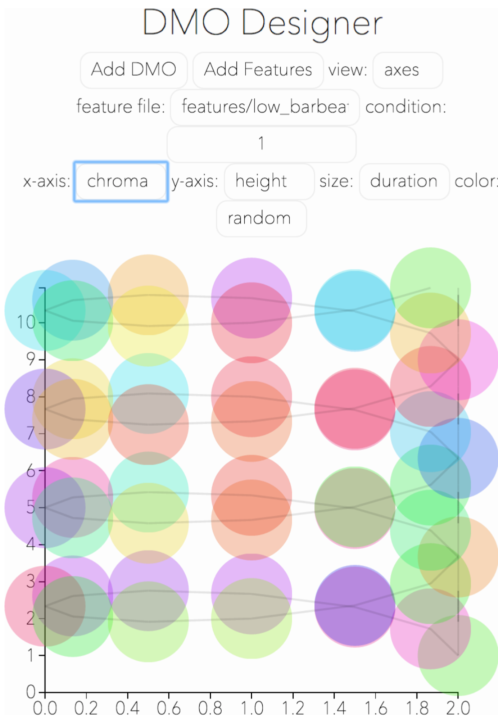

Figure 2. A Dynamic Music Object rendered as a pitch helix.

parameters of the player. Figure 2 shows the helical structure when created in the DMO Designer. In addition to these mappings, we also map the compass and accelerometer controls to the listener's position in space, which allows them to move around within the helical structure.

4. CONCLUSION AND FURTHER WORK

The approach we demonstrate in this simple example can easily be generalized by highlighting other semantic characteristics of an audio recording, such as dynamics, instrumentation, or expressive content, or by mapping to parameters other than spatial position.

5. REFERENCES

[1] Simon Dixon, Werner Goebl, and Gerhard Widmer. The" air worm": An interface for real-time manipulation of expressive music performance. *Proceedings of the International Computer Music Conference (ICMC)*, pages 614–617, 2005.

[2] Sebastian Ewert, Bryan Pardo, Meinard Müller, and Mark D. Plumbley. Score-informed source separation for musical audio recordings: An overview. *IEEE Signal Processing Magazine*, 31(3):116–124, May 2014.

[3] Roger N. Shepard. Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89(4):305, 1982.

[4] Florian Thalmann and Guerino Mazzola. Visualization and transformation in general musical and music-theoretical spaces. In *Proceedings of the Music Encoding Conference 2013*, Mainz, 2013. MEI.