AniMorph: Animation Driven Audio Mosaicing

Augoustinos Tsiros and Grégory Leplatre
Centre for Interaction Design
Edinburgh Napier University
10 Colinton Road, EH10 5DT, UK
a.tsiros@napier.ac.uk, g.leplatre@napier.ac.uk

1. INTRODUCTION

This paper describes *AniMorph* a system for animation driven Concatenative Sound Synthesis (CSS). We can distinguish between two main application domains of CSS in the context of music technology: target sound re-synthesis, and free sound synthesis. The difference between these two categories is that in target sound re-synthesis the aim is to re-create a sound or a sound's characteristics by providing audio examples (see Schwarz & Schnell 2010, Stevens et al. 2012), while free sound synthesis focuses on exploration of the audio corpus in order to synthesise novel sounds that do not necessarily resemble the features of another sound, for examples see (Comajuncosas 2011, Navab et al. 2014, Schwarz & Hackbarth 2012).

The main motivation for the present investigation is to (i) develop appropriate models of interaction for efficient exploration of the audio corpus, and (ii) develop perceptually meaningful mappings to enable practitioners to create novel sounds using CSS by specify the perceptual characteristic of the sound that they want to synthesise in visual terms. The present research considers that it is of paramount importance to achieve an intuitive mapping in order to enable interaction with concatenative synthesis for creative purposes (e.g. sound design, electroacoustic composition, live performance). *AniMorph* builds on the software developed through an earlier system: *Morpheme* that uses sketching as a model for interaction (Tsiros 2013). To expand upon this work, we modified the existing interface in order to make it work with animation as user input.

2. SYSTEM OVERVIEW

AniMorph has been developed using the Max/MSP visual programming environment. *AniMorph* provides image processing capabilities (i.e., visual effects) that can be controlled by the user using touch-based interface (i.e., using a multi-touch screen).

Additionally, the system supports connectivity with third party software applications that have built-in image processing and animation capabilities (e.g., VJ, image processing and animation software packages) and support the Open Sound Control (OSC) protocol.

The architecture of the system is presented in Figure 1. *AniMorph* performs statistical analysis on the visual input in order to extract a stream of visual descriptors; for more details about visual feature extraction see Figure 2. In the present version of *AniMorph* the analysis of the visual input results in a four dimensional feature vectors which is used as the target for querying audio-units from the database of the *CataRT* system (Schwarz 2004). The *CataRT* system performs a nearest neighbour search to find the best match between the target vector and the audio descriptors in the database. Sound synthesis is accomplished by combining the audio-units that are retrieved from the database, creating a new sound sequence. This approach to sound synthesis is known as audio mosaicing. Table 1 shows the current mapping for the retrieval of audio-units from the database. Video examples that demonstrate the *AniMorph* system in action are available online:

<https://avrenderstudy.wordpress.com>

3. FUTURE WORK

The present paper proposes a system for animation driven corpus based concatenative synthesis. The next step in this research will be to increase the audio-visual associations of the mapping and further refine/optimize the mapping between audio and visual feature through empirical testing. Deploy more elaborate methods (i.e. pattern recognition and classification) in order to extract high level visual features from the animation. Finally, *AniMorph* can easily be re-appropriated to sonify visual scenes as a sensory substitution system.

Figure 1: AniMorph's system overview

Figure 2: Visual feature extraction. RGB stands for red, green, blue and HSL for hue, saturation and lightness

Table 1: Mapping between audio and visual descriptors

Audio Features	Visual Features
Spectral flatness	Texture compactness
Spectral Centroid	Brightness
Periodicity	Texture Entropy
Loudness	Saturation

4.0 REFERENCES

- Comajuncosas, J. (2011) Nuvolet: 3d gesture-driven collaborative audio mosaicing. *New Interfaces for Musical Expression*.
- Navab, N., Nort, D. Van, & Wei, S. (2014) A Material Computation Perspective on Audio Mosaicing and Gestural Conditioning. *New Interfaces for Musical Expression*.
- Schwarz, D. (2004) *Data-driven concatenative sound synthesis*. Universite Paris 6, France.
- Schwarz, D. & Hackbarth, B. (2012) Navigating variation: composing for audio mosaicing. *International Computer Music Conference*.
- Schwarz, D. & Schnell, N. (2010) Descriptor-based sound texture sampling. In *Sound and Music Computing Conference*.
- Stevens, S. A., Doug, L. J., & Marschner, S. (2012) Motion-driven concatenative synthesis of cloth sounds. *ACM Transactions on Graphics*, 31(4), 1–10.
- Tsiros, A. (2013) A Multidimensional Sketching Interface for Corpus Based Concatenative Synthesis. In *International Conference on Auditory Display*.

Figure 3: Example of three audio files that were synthesised from the images using the same instrumental audio corpus