Timbre Morphing: Near Real-time Hybrid Synthesis in a Musical Installation

Duncan Williams
Interdisciplinary Centre for Computer Music Research, Drake Circus, Plymouth University, Devon, PL4 8AA, United Kingdom
Duncan.williams@plymouth.ac.uk

Peter Randall-Page
P.O. Box 5
Drewsteignton, Devon, EX6 6QN,
United Kingdom

Eduardo R. Miranda
Interdisciplinary Centre for Computer Music Research, Drake Circus, Plymouth University, Devon, PL4 8AA, United Kingdom
Eduardo.miranda@plymouth.ac.uk

ABSTRACT

This paper presents an implementation of a near real-time timbre morphing signal processing system, designed to facilitate an element of 'liveness' and unpredictability in a musical installation. The timbre morpher is a hybrid analysis and synthesis technique based on Spectral Modeling Synthesis (an additive and subtractive modeling technique). The musical installation forms an interactive soundtrack in response to the series of Rosso Luana marble sculptures *Shapes in the Clouds, I, II, III, IV & V* by artist Peter Randall-Page, exhibited at the Peninsula Arts Gallery in Devon, UK, from 1 February to 29 March 2014.

The timbre morphing system is used to transform live input captured at each sculpture with a discrete microphone array, by morphing towards noisy source signals that have been associated with each sculpture as part of a pre-determined musical structure. The resulting morphed audio is then fed-back to the gallery via a five-channel speaker array. Visitors are encouraged to walk freely through the installation and interact with the sound world, creating unique audio morphs based on their own movements, voices, and incidental sounds.

Keywords
Timbre morphing, surround installation, live signal processing

1. INTRODUCTION

Timbre morphing is a relatively recent signal processing technique that can be approached in a variety of ways [1]–[5]. In order to fully explain the principles behind the system of timbre morphing used in this paper an understanding of the distinction between timbre morphing and general audio morphing (also referred to as acoustic morphing) is necessary.

Audio morphing is a technique for creating a hybrid sound with characteristics derived from both a source and target sound. In an audio morpher, a source and target sound are analysed, and a specification of some form of acoustic middle ground is then determined. Typically, this specification is used as a feature set from which a hybrid signal is synthesized. Audio morphing has applications in speech and singing voice manipulation, broadcast and DJ playback, and novel sound design.

By contrast, a *timbre* morpher is a system for audio morphing that is adapted to allow the user to vary individual timbral characteristics (or *timbral attributes*) between those of the source and target sounds. A full timbre morpher should allow the user to create a new source feature set with independent control over a range of discrete timbral attributes without affecting unintended timbral or acoustic characteristics from the source sound.

In a simple audio morph, the source (A), becomes the target (B). *All* acoustic and timbral attributes are morphed. A 50% morph would constitute a 6 dB attenuation before summing both source and target sound. In a split morph, some of the attributes are morphed, either through independent acoustic, or timbral control. This distinction is further illustrated in Figure 1.
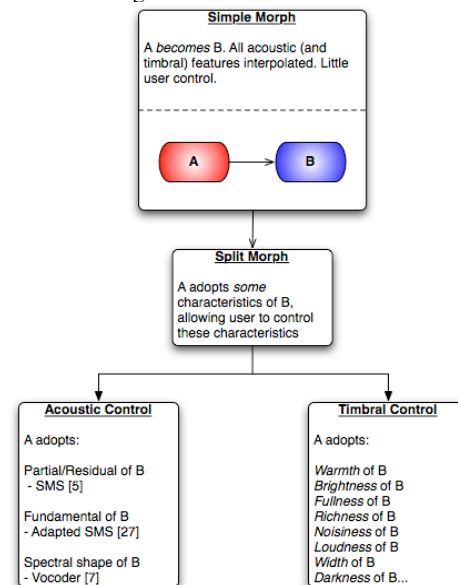

Figure 1. Family tree illustrating different types of morph and the characteristics over which they offer control

A prototype timbre morpher has been previously designed, developed, and evaluated by listener testing and statistical analysis [6]–[8]. This morpher is based on an existing audio morpher which uses Spectral Modeling Synthesis (SMS) [9], [10] to analyse the source and target sounds, interpolate their characteristics, and synthesise a hybrid output. Various computer implementations of cross-synthesis, audio morphing, and sound hybridization were investigated in order to inform the selection of a suitable open-source platform on which to base the prototype timbre morpher. SMS was the most suitable of the available alternatives, partly as a result of its fast processing speed due to its use of a stochastic process (rather than multiple oscillators) to handle non-discrete spectral elements. A routine for adapting the existing SMS audio morpher to timbral control was proposed whereby specific modules are incorporated for the extraction and interpolation of the known acoustic correlates of a selected timbral attribute. A series of listening tests involving listener evaluation of a hybrid stimulus set created by the prototype morpher demonstrated that it was capable of morphing the timbre of a source sound toward that of a target sound specifically in terms of its

brightness, *softness*, and/or *warmth*, independently from its other perceptual attributes such as *richness*, *noisiness*, *punch*, et al. Ultimately, this provides a novel tool for creative sound design and composition.

This paper describes an implementation of this timbre morpher as a compositional tool in the creation of a new musical installation, *Concord for Five Elements*, commissioned by the Peninsula Arts Contemporary Music Festival and first performed at the opening of Peter Randall-Page's *New Sculpture and Works on Paper*, in response to his new sculpture series in Rosso Luana marble *Shapes in the Clouds, I, II, III, IV & V*, which were exhibited at the Peninsula Arts Gallery from 1 February to 29 March 2014. These sculptures provided the seed material for the musical structure of the installation.

2. SYSTEM OVERVIEW

Live input from one of five small diaphragm condenser microphones in the gallery is analysed by FFT for spectral and temporal characteristics. These are used to generate control signals for timbral attribute selection and to specify the amount of morphing required. The live input is also used to generate pitch vector and noise residuals as per a normal SMS analysis, in which a peak picking phase selects pitch content to store as a vector that is then subtracted from the input data to give a 'noisy' residual. The pitch vector and noise residual are then stored as target data. The source data for the morph is a pre-existing block of residual and pitch vector data that is then hybridized towards the target data according to the extracted control signals. The hybridized pitch and residual data is then summed and synthesized by IFFT, before being routed to one of the loudspeakers in the array. An overview of the timbre morphing system is shown in Figure 2.

With any multi-attribute system, there is a possibility of introducing unintended timbral change when manipulating a single attribute, particularly if the attributes in question have some overlapping acoustic correlations. Three ways of tackling any undesirable timbral change that might occur as a by-product of the morphing were devised.

Firstly, an automatic system was coded, whereby an inverse morph of the overlapping attribute would be applied to the output, returning the hybrid signal to the source value in the unintended timbral attribute. Secondly, a lookup table of acoustic correlates and their corresponding impact on each timbral attribute was considered. In this instance, attributes with some unique acoustic correlation could then be compensated with an increase or decrease in the appropriate correlate as necessary, though this would limit possible attribute selection in the future. Thirdly, a non-correcting 'display-only' system was suggested, such that the user-interface would simply indicate that if attribute *a* was adjusted by *x*% then attribute *b* would alter by *y*% in sympathy. The routines for discrete morphing of various specific timbral attributes was previously evaluated in a series of perceptual listening tests, quantifying the difference between morphs of different attributes by means of perceptual scaling with a multi-dimensional scaling analysis in <4D, and quantifying perceived changes with a verbal elicitation experiment and protocol analysis. The final prototype morpher employs a combination of these techniques, but for full details of the underlying timbral compensation approaches taken with each attribute, the reader is referred to [6-8].
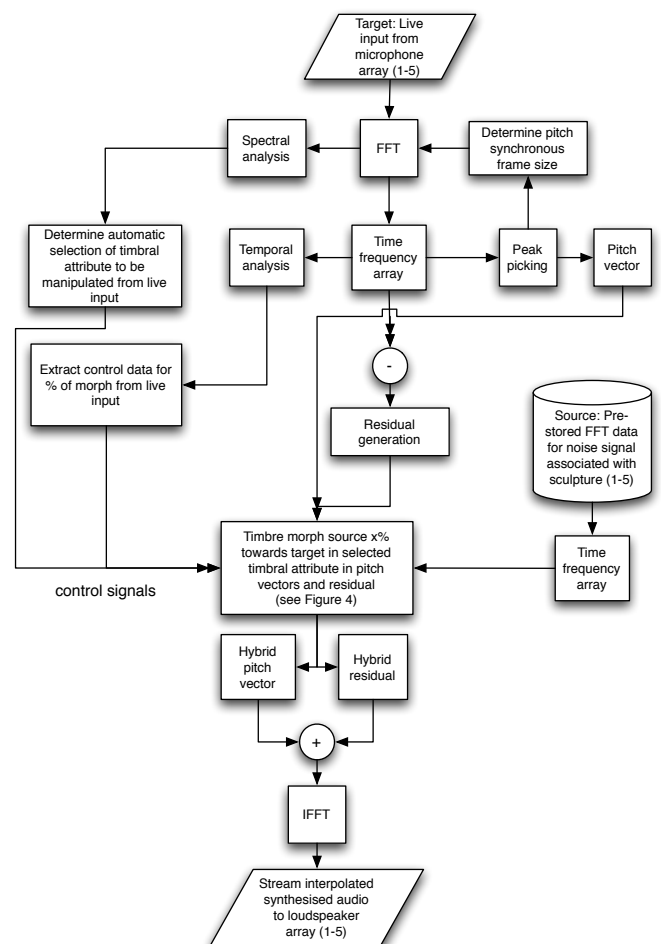

Figure 2. Overview of system. A pre-stored array of pitch vector and residual data is morphed towards pitch and residual data from a live microphone input analysed by FFT. The window size of the FFT carried out on the live microphone feeds is pitch-synchronous to improve time/frequency resolution.

3. SCULPTURES AND SOURCE SOUNDS

If these giant shapes could talk, what would they say to one another? How different would their voices be? And, how might they react with the audience in this fixed installation? Can the shapes develop a shared prosody with their visitors by using their own voices to mirror the sounds around them? These are the driving inspirations behind the installation in *Concord for Five Elements*.

In *Concord for Five Elements*, Peter Randall-Page's five "Shapes in the Clouds" are accompanied by a spatialised speaker array creating an evolving soundfield within which the audience is encouraged to move freely. Inside this soundfield, each of the sculptures is represented by a unique acoustic voice that is created by the near real-time timbre morphing routine, morphing between the live target signal, and a series of existing 'noisy' source signal.

The sculptures themselves are based on the five platonic solids (see Figure 3), hence, each sculpture is initially voiced by a noisy sound derived from Plato's *Timaeus* (*fire* with the tetrahedron, *earth* the cube, *air* the octahedron, *space* the dodecahedron, and *water* the icosahedron).

Figure 3. Three of the "Shapes in the Clouds" sculptures, showing differing arrangements of platonic solids.

Source material for fire, air, and water was recorded in and around Dartmoor, the home of Randall-Page's studio. Earth was sourced from another location recording of the crunching of footsteps on the moor. Perhaps (to modern ears at least) the most cryptic element, Space, had a source derived from the reverberant properties of each of the other elements combined. These sounds were edited, processed, and looped to derive the main source signals for the installation. The timbre morphing routine is flexible enough that it can either run totally independently or with a degree of user control. In the case of *Concord for Five Elements*, some compositional structure was imposed by fading noisy source signals in and out at given points in the overall timeline. Thus, a combination of pre-processed and live sounds were used (with a fixed buffer of 2000ms used to handle the associated processing lag incurred by the live morphing).

4. INSTALLATION, TARGET SOUNDS, AND HYBRID OUTPUTS

Figure 4 shows the positioning of the sculptures and loudspeaker array in the installation at the Peninsula Arts Gallery, Devon, UK. Small diaphragm capacitor microphones were placed above each sculpture and fed directly to an 8in, 8out audio interface. Each microphone feed was used to provide the target signal for a channel of timbre morphing, the output of which was then routed to the corresponding loudspeakers. The morphed sounds themselves are played back via the multichannel loudspeaker array as a contributory part of the overall musical structure – some linear progression is provided by additional synthesized instrumentation and percussion voices.

As shown in Figure 2, the amount and type of timbre morph is controlled by the spectral and temporal properties of the input (target) signal. Signals with high rhythmic density and large amplitudes result in greater degrees of morphing. Target signals with larger spectral flux and higher spectral centroid result in brightness and warmth morphs. Target signals with low spectral flux trigger morphs in the temporal domain (noisiness and hardness). Figures 6-8 show the progress of an example morph.
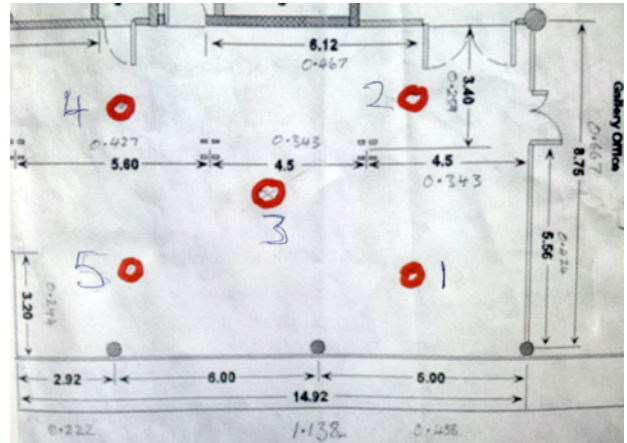

Figure 4. Position of sculptures and corresponding loudspeaker array in Peninsula Arts Gallery. The numbers correspond with platonic structures and their associated elements in *Timaeus* as follows: 1 = the dodecahedron (space), 2 = the tetrahedron (fire), 3 = the cube (earth), 4 = the octahedron (air) and 5 = the icosahedron (water)
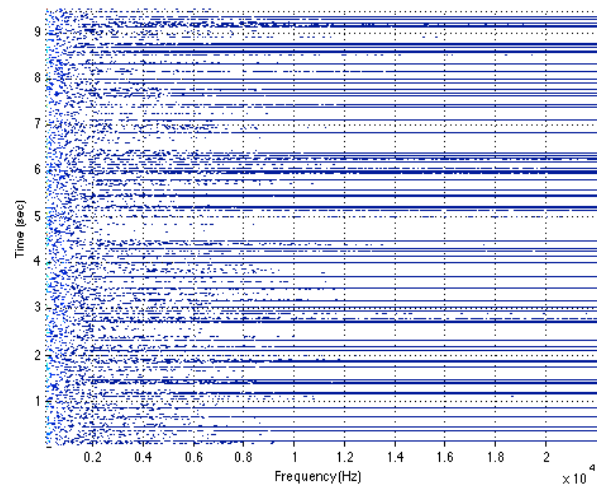

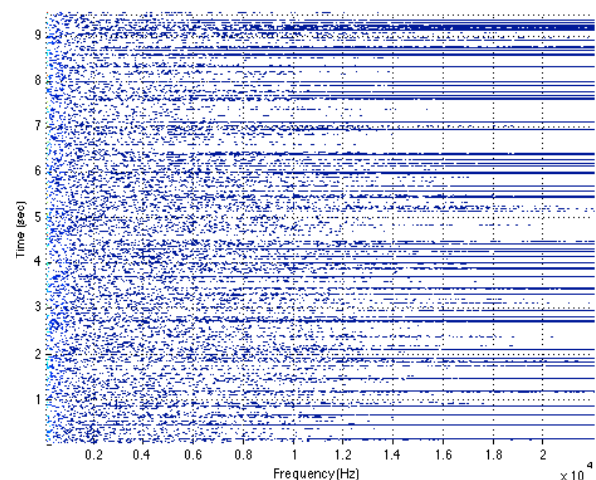**Figure 6. Spectrogram of a short sample live input as a target.**

Figure 7. Spectrogram of an accompanying noise profile used as a source sound in sample timbre morph.
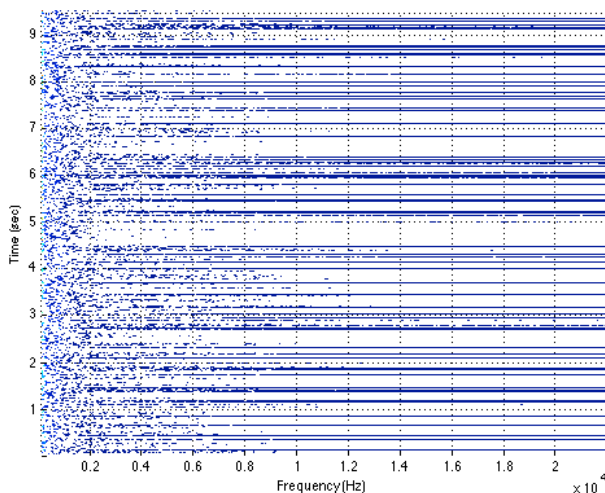
Figure 8. Spectrogram of a finished timbre morph with a control signal of 50% in warmth attribute. Regions with strong harmonic content are preserved with addition of noisy partials, whilst an overall spectral tilt is applied to neutralize subsequent changes in spectral centroid.

The morphed output shown in Figure 9 maintains some of the temporal information from the target signal, whilst most of the harmonic content from the source signal remains unaffected by the processing. In practice, the sonic output of the timbre morpher varies quite radically depending on the amount and type of perceptual attribute which is being manipulated, and of course on the quality of the target signal captured from the microphone array. For example, the movement of a noisy source such as the *air* recordings (wind on Dartmoor) towards a target sound of footsteps on the polished concrete floor within the reverberant environment of the gallery yields morphs with thick, percussive textures. Movement towards speech sounds, as occasionally happens when gallery visitors make comments or engage with the marble pieces, creates unusual vocoder-like effects, though the results are often less obvious than those produced by a traditional vocoding process. The overall combination creates a sound-world which gives the audience some degree of familiarity if it is absorbed unconsciously, but some visitors reported that when the sounds are listened to more carefully, the resulting timbres are quite alien.

5. CONCLUSIONS AND FURTHER WORK

A novel signal processing technique which falls under the umbrella of *timbre morphing* has been implemented using a multi-input microphone system to provide the target sound for near real-time live morphing in a musical installation. This is, to the best of the author's knowledge, the first time such a system has been used in a real-time musical installation, though timbre morphing has been adopted as a theoretical term, and prototyped in the signal processing domain by a number of researchers in the last fifteen or so years. In the case of *Concord for Five Elements*, the timbre morphing routine was used to respond to, and in a sense create an interactive sonification of, the large marble sculptures. The timbre morpher allows for the creation of sounds which are novel yet retain an element of interactivity with the visitors (who are also allowed to touch the shapes). Whilst an objective evaluation of the success of the installation is difficult, some types of timbre morph were seemingly more aesthetically successful than others, with strong percussive sounds being generated by various combinations of noisy sources and transient target

signals. A possible refinement to create more congruent outputs across the range of morph types might be to daisy-chain the output of these processes in the case of target sounds comprised of speech, such that the sonified output still retains a sense of the scale and composition of the physical objects in the installation.

In future it would be useful to expand this timbre morpher by the inclusion of additional attributes – the current prototype only offers a limited number of timbral controls. Computationally, the timbre morpher is fairly slow, at approximately 3-6% of real time (a 10 second hybrid typically takes between 0.3-0.6 seconds to process, but can take longer depending on the amount of morphing and the complexity of the input signals, hence a 2000ms buffer was used in this installation to handle the corresponding processing lag). By combining streamlined code, in particular the iterative calculations involved when determining neutral spectral centroid values as part of the brightness morphing routine, the platform should ultimately aspire to totally real-time operation when combined with look ahead and faster machine processing.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1]   G. Cospito and R. de Tintis, "Morph: Timbre Hybridization tools basedon frequency," in Workshop on Digital Audio Effects (DAFx-98), 1998.

[2]   L. Haken, K. Fitz, and P. Christensen, "Beyond traditional sampling synthesis: Real-time timbre morphing using additive synthesis," in Analysis, Synthesis, and Perception of Musical Sounds, Springer, 2007, pp. 122–144.

[3]   C. Hope and D. Furlong, "Time-Frequency Distributions for Timbre Morphing: The Wigner Distribution Versus the STFT," in Procceedings of the SBCMIV, (4th Symposium of Brasilian Computer Music), Brasillia, 1997.

[4]   N. Osaka, "Timbre morphing and interpolation based on a sinusoidal model," The Journal of the Acoustical Society of America, vol. 103, p. 2757, 1998.

[5]   E. Tellman, L. Haken, and B. Holloway, "Timbre morphing of sounds with unequal numbers of features," Journal of the Audio Engineering Society, vol. 43, no. 9, pp. 678–689, 1995.

[6]   D. Williams, T. Brookes, "Perceptually-Motivated Audio Morphing: Brightness," in Audio Engineering Society Convention 122, 2007.

[7]   D. Williams, T. Brookes, "Perceptually-motivated Audio Morphing: Softness," in Proceedings of the 126th Audio Engineering Society Convention, Munich, 2009, p. Paper Number 7778.

[8]   D. Williams, T. Brookes, "Perceptually-Motivated Audio Morphing: Warmth," in Audio Engineering Society Convention 128, 2010.

[9]   X. Serra and J. I. Smith, "Spectral Modelling Synthesis," Computer Music Journal, vol. 14, no. 4, 1990.

[10]   X. Serra and J. Bonada, "Sound Transformations Based on the SMS High Level Attributes," in International Conference on Digital Audio Effects, Barcelona, 1998.