Timbre interfaces using adjectives and adverbs

Colin G. Johnson
Computing Laboratory
University of Kent
Canterbury, England
C.G.Johnson@kent.ac.uk

Alex Gounaropoulos
Computing Laboratory
University of Kent
Canterbury, England
ag84@kent.ac.uk

ABSTRACT

How can we provide interfaces to synthesis algorithms that will allow us to manipulate timbre directly, using the same timbre-words that are used by human musicians to communicate about timbre? This paper describes ongoing work that uses machine learning methods (principally genetic algorithms and neural networks) to learn (1) to recognise timbral characteristics of sound and (2) to adjust timbral characteristics of existing synthesized sounds.

Keywords

timbre; natural language; neural networks

1. INTRODUCTION

When human musicians "interface" with other human musicians, we do so in words. One way in which we use words is to communicate about timbre: we say "can you make the sound shine more", "I'd like to get a really gritty sound", "let's try to play that more warmly".

This ability to use natural language descriptions of timbre (whether "straight" or metaphorical) is typically absent from interfaces with music technology devices. As a result, users have to either have a very strong understanding of the underlying mechanisms that produce the sound, or a large amount of "trial-and-error" experience with generating timbral changes within a system [5, 11]. A small number of attempts [2, 7] have attempted to create systems that offer an intuitive interface to timbre; however, there appears to be little recent work in this area [9].

By *timbre* we will mean the micro-level spectral characteristics of sound as discussed by Wishart [10], as opposed to the gross timbral distinctions [6] used e.g. in the MPEG-7 standard.

In this paper we discuss ongoing work which applies machine learning methods to associate changes in synthesis parameters with changes in timbre described by adverbs and adjectives. Overall the system is structured as illustrated in figure 1. Before the system is used, a training process is carried out whereby the system learns to associate features of sounds with certain timbre-description words that have been allocated by a human listener. Then, when the user wants to adjust the sound with regard to a particular word, an algorithm that adjusts the synthesis parameters (e.g. with a genetic algorithm), uses the synthesis algorithm to generate a sound with regard to that parameter-set, then uses the trained classifier to measure the timbral effects. The levels of these timbral effects are then measured and fed as a *fitness measure* into the parameter adjustment algorithm.

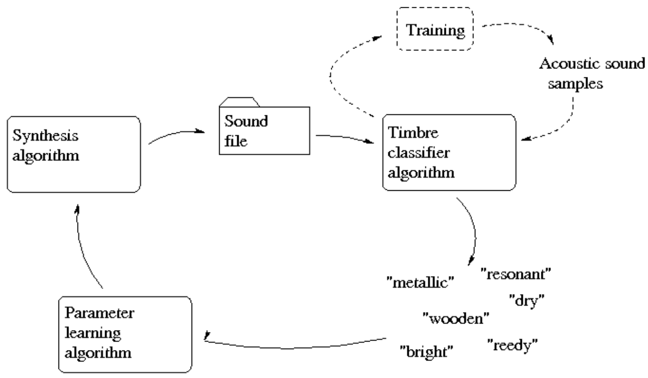

Figure 1: Indirect modification of timbres.

The remainder of the paper describes the two main parts of the system: the timbre classification program, and the parameter learning algorithm. Full details of this can be found in our paper [3].

2. PART 1: TIMBRE RECOGNITION ALGORITHM

Timbre recognition can be seen as a *classification* problem of the type well studied in machine learning [8]. In such a problem we have a number of data items each of which is described by the values of a number of *attributes*; each data item also fits into one or more *classes*. In this case each data item is a sound. The attributes of that sound are a number of measures that are made on that sound: the relative peak amplitude of the first 15 partials of the sound, the detuning of the partials, and measures of the amplitude envelope acquired by fitting an ADSR envelope to the sound. The classes are a number of timbre-words (drawn mostly from the list in [2]); a human listener has, for each sound, rated how much they believe that that word describes the timbral characteristics of the sound. In our current program, the words used are *bright, warm, harsh, hit, plucked, constant, thick, metallic,* and *woody.*

A common approach to such classification problems is

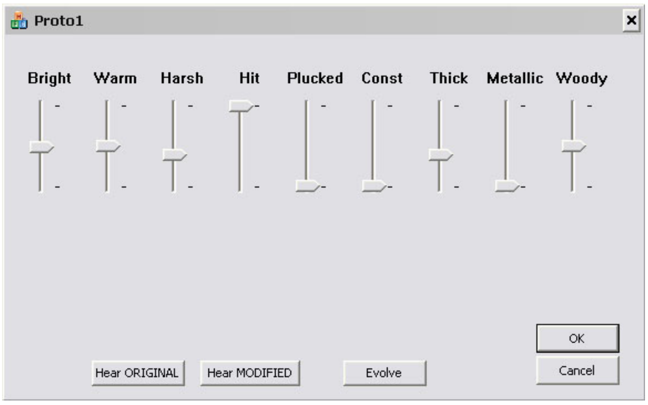

Figure 2: Adjusting timbres with sliders.

to use neural networks to learn which patterns in the attributes map on to which classes. This is the approach that we have used here. The inputs to the network are the attribute values described above, the outputs the timbre words. We use the well-known *back-propogation* algorithm (e.g. [1]) to learn the network weights using a large number of training examples. To test the classification, we fed a number of (attributes describing) new sounds, not used in training, through the network, and observed the value given for each of the timbre words by the network. Overall, a good classification of timbres was obtained when compared to classification by a human listener (details in [3]).

3. PART 2: PARAMETER ADJUSTMENT ALGORITHM

The second main part of the system is an algorithm that adjusts the parameters of the synthesis algorithm to respond to a desired timbre change. The current interface to the system is illustrated in figure 2. The user inputs a sound, the neural network assesses the timbral characteristics of the sound according to the ten timbre-words used, and the system sets the sliders accordingly.

The user then makes changes to the sliders to reflect the desired change. We have experimented with various algorithms to generate the desired change in synthesis parameters. In the first, we use a genetic algorithm to find a new setting for the synthesis parameters that (1) is close to the original settings (so that the overall sound does not change) and (2) reflects the desired timbral change. A second approach involves feeding the new desired values for the timbral characteristics through the neural network, to find regions of parameter-space that are associated with these settings, then finding the smallest parameter-space move required to go from the original sound to those regions that produce the desired timbre. Both approaches produce decent results; at present the second approach seems more promising.

4. RESULTS

Some example sounds can be heard at `http://www.cs.kent.ac.uk/people/staff/cgj/research/nime2006/nime2006.html`. We will demo the system live at the conference.

5. ONGOING WORK

We are currently pursuing a number of future directions for this work:

- Improved timbre recognition, for example by using ear-like pre-processing steps [4].
- Focusing more effort on having the system *not* adjust those characteristics of the sound that are not relevant to the current
- Interfacing this to e.g. MIDI control wheels to allow on-the-fly live manipulation of timbre.
- Creating systems that learn in advance the directions in parameter space which affect a particular timbre change, rather than running the parameter-adjustment algorithms on the fly. This may be a prerequisite for the use of the system in a live environment.

6. REFERENCES

[1] R. Beale and T. Jackson. *Neural Computing.* Institute of Physics, 1990.

[2] R. Etherington and B. Punch. SeaWave: A system for musical timbre description. *Computer Music Journal*, 18(1):30–39, 1994.

[3] A. Gounaropolous and C. Johnson. Synthesising timbres and timbre-changes from adjectives/adverbs. In F. Rothlauf et al., editor, *Applications of Evolutionary Computing.* Springer, 2006.

[4] D. M. Howard and A. M. Tyrrell. Psychoacoustically informed spectrography and timbre. *Organised Sound*, 2(2):65–76, 1997.

[5] C. G. Johnson. Exploring the sound-space of synthesis algorithms using interactive genetic algorithms. In G. A. Wiggins, editor, *Proceedings of the AISB Workshop on Artificial Intelligence and Musical Creativity, Edinburgh*, 1999.

[6] S. McAdams, S. Winsberg, S. Donnadieu, G. de Soete, and J. Krimphoff. Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58:177–192, 1995.

[7] E. R. Miranda. An artificial intelligence approach to sound design. *Computer Music Journal*, 19(2):59–75, 1995.

[8] T. Mitchell. *Machine Learning.* McGraw-Hill, 1997.

[9] A. Seago, S. Holland, and P. Mulholland. A critical analysis of synthesizer user interfaces for timbre. In *Proceedings of the XVIIIrd British HCI Group Annual Conference.* Springer, 2004.

[10] T. Wishart. *Audible Design.* Orpheus the Pantomime, 1994.

[11] T. Wishart. *On Sonic Art.* Harwood Academic Publishers, 1996. second edition, revised by Simon Emmerson; first edition 1985.