Timbre Similarity: An overview

By Daniel McEnnis

Timbre similarity is a massive area with a large number of important papers in the past 30 years. The first of these papers was by Grey and Gordon in 1978 (Grey and Gordon 1977). Early work (prior to the 90's) was focused primarily on evaluating perceived similarity of listeners. Later work had a focus on computers classifying timbres (Brown 1999, Fujinaga 1998, Brown 1999, Fujinaga and MacMillan 2000, Herrera et al. 2002, Tindale et al. 2004). While not a complete overview, a handful of papers are presented to give an idea of the nature of the field.

Grey and Gordon was the first paper to be published on timbre similarity (Grey and Gordon 1978). This paper utilized 16 instrument sounds taken from a 14 different instruments. Each sound had three simplifications that were also added to the experiment. The study had listeners listen to every possible pairing of the sample notes and rate each pair based on its similarity. Afterwards, Grey and Gordon utilized a method known as multidimensional scaling (MDS) to determine the number of dimensions of timbre. MDS accomplishes this by attempting to place the results in an Euclidian space and determining the minimum number of dimensions required to maintain the relationships expressed by the data (within some noise). This process found that 3 dimensions were necessary to explain the data.

Later work by Fujinaga (Fujinaga 1998) centered on creating systems for automatically classifying musical sounds into a taxonomy. This system worked exclusively on the steady state of the tone, hand edited from McGill Instrument Samples cds. Fujinaga used the first ten moments, fundamental frequency, and amplitude as features. To classify the instrument sounds, Fujinaga used a genetic algorithm to filter the data and a K Nearest Neighbor (KNN) classifier to classify the results. Fujinaga acquired a 64% classification rate across 39 instruments with this methodology.

Judith Brown also did a study on instrument classification, though using a very different technique (Brown 99). She utilized excerpts from solo recordings of saxophone and oboe as data for her experiment. It is interesting to note that a single length sound example of oboe and saxophone each was used as the training sample, with all other samples used as testing data. She extracted mel frequency cepstrum coefficients (MFCCs) as her features and utilized a Gaussian classifier in her experiment.

Fujinaga and MacMillan (Fujinaga and MacMillan 2000) extended Fujinaga's original 1998 paper, making a realtime instrument recognition system with feature extraction implemented in PD. It used a superset of Tujunga's features, adding velocity of the spectral centroid, spectral irregularity, and tristimulus as features. The classifier used was genetic algorithms and KNN. The results improved from 64% to 68%.

Herrera, Yeterian, and Gouyon implemented a drum classifier that classifies a variety of different drum sounds (Herrera et al. 2002). The data set was 634 drum sounds pulled from different commercial cds of sample sounds. The feature set was divided into attack and decay features with a great deal of overlap between the two sets. They also extracted 8 empirically chosen bands and 13 MFCCs. They tested a KNN, canonical discriminant analysis, and C4.5 classifier on the data set and received results of 99.2% for KNN, 97.2% for C4.5, and 99.1% for canonical discriminant analysis.

Tindale, Kapur, Tzanetakis, and Fujinaga created another classifier for classifying the different types of snare drum hits. They used spectral flux, spectral rolloff, spectral centroid, spectral kurtosis, skewness, 12 MFCCs, linear prediction coding coefficients, and 9 wavelet bands with a mean and variance for each band. The classifiers used were KNN, Support Vector Machines, and a neural net. KNN recorded a 94.9% success rate, the support vector machine recorded a 68.8% success rate, and the neural net recorded an 89% success rate.

While not a complete overview, these papers give a taste of the current field of timbre similarity research.

Brown, J. 1999. Computer identification of musical instruments using pattern recognition with cepstral coefficients as features. *Journal of the Acoustical Society of America*. 105: 1933–41.

Fujinaga, I. 1998. Machine recognition of timbre using steady-state tone of acoustic musical instruments. *Proceedings of the International Computer Music Conference*. 207–10.

Fujinaga, I. and K. MacMillan. 2000. Realtime recognition of orchestral instruments. *Proceedings of the International Computer Music Conference*. 141–3.

Grey, J., and G. Gordon. 1978. Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America.* 63(5): 1493–1500

Herrera, P., A. Yeterian, and F. Gouyon. 2002. Automatic classification of drum sounds: A comparison of feature selection methods and classification techniques. International *Conference on Music and Artificial Intelligence*. 2: 69–80.

Tindale, A., A. Kapur, G. Tzanetakis, and I. Fujinaga. 2004. Retrieval of percussion gestures using timbre classification techniques. *Proceedings of the International Conference on Music Information Retrieval*. 541–5.