Classification of Timbre Similarity

Corey Kereliuk
McGill University

March 15, 2007

Before discussing techniques to assess the similarity of different timbres, it would be prudent to review the definition of *timbre*. In fact, the concept of timbre has been difficult to elucidate. Timbre is often described, not in terms of what it *is*, but rather what it is *not*. The Oxford English Dictionary defines timbre as:

> The character or quality of a musical or vocal sound (distinct from its pitch and intensity) depending upon the particular voice or instrument producing it, and distinguishing it from sounds proceeding from other sources

Wessel (1979) says, "Timbre refers to the 'color' or quality of sounds, and is typically divorced conceptually from pitch and loudness". Wessel, however, goes further in his definition, saying,

> Perceptual research on timbre has demonstrated that the spectral energy distribution and temporal variation in this distribution provide the acoustical determinants of our perception of sound quality

Wessel and Grey (1975) conducted several experiments in order to illuminate the perceptual relevance of timbre. In these experiments participants were made to listen to several sound samples, one pair at a time. They were asked to assess, on a scale from 0-9, the similarity between each set of samples. Using this data a *dissimilarity* matrix was constructed. A multidimensional scaling algorithm (MDS) was then used to reduce the dimensionality of the the data to a two-dimensional *timbre space*. The distance between instruments in the space was inversely proportional to their similarity.

There are many reasons to study tools to extract the similarity of different timbres, including:

- Psychoacoustic studies
- Musicological analyses
- Source separation
- Instrument identification

- Content-based management systems for the navigation of large catalogues
- Composition
- Identifying bird calls from the same species
- Speaker identification
- etc.

There are several important considerations for timbre recognition systems. For one, it must be decided whether monophonic, or polyphonic timbre will be recognized. The former problem is well understood, but the latter is still unsolved, and an area of active research. Timbre recognition can also be done on a local scale, looking at fine variations in the micro-structure of a signal, or at the global scale, looking at long-term statistics. It should also be noted whether the chosen scheme produces perceptually relevant results.

Common features used in timbral analysis are Mel-Frequency Cepstrum Coefficients (MFCCs), Spectral Centroid, Log-attack-time, and Spectral Flatness (Degree of noisy-ness). A variety of statistical classification techniques are also used including Principle Component Analysis (PCA), Gaussian Mixture Models (GMMs), Hidden Markov Models (HMMs), Genetic Algorithms (GAs), and Neural Networks (NNs) (Herrera-Boyer, Peeters, and Dubnov, 2003).

In this summary I examine global polyphonic timbre description, which uses long-term statistics to determine timbral similarity. The essential idea behind this technique is to average a set of features over the duration of the signal (For example MFCCs (Aucouturier, Pachet, and Sandler, 2005)). One might expect the result to be flat or noisy, however, it turns out that a *global* shape emerges, which tends to be quite specific to a given texture. Figure 1 illustrates the concept of a global spectral envelope. This envelope was created by averaging MFCCs from 500, 50ms frames. Aucouturier (2005) proposes modeling the MFCCs as a mixture of Gaussians.

$$p(F_t) = \sum_{m=1}^{M} \pi_m \aleph(F_t, \mu_m, \Gamma_m)$$

Here the feature vector F_t at time t is modeled as the sum of M Gaussians with mean μ_m and variance Γ_m. The GMM is initialized by k-mean clustering and trained using the classic EM algorithm. Figure 2 shows a GMM with M=3 (the dots are MFCCs). In order to compare the timbral similarity of two songs using the GMM a number of sampling points are chosen from each song, and then the following similarity measure is computed:

$$D(A,B) = \sum_{i=1}^{N} logP(S_i^A|A) + \sum_{i=1}^{N} logP(S_i^B|B) - \sum_{i=1}^{N} logP(S_i^A|B) - \sum_{i=1}^{N} logP(S_i^B|A)$$

where N is the number of sampling points used and S_i^A is the i^{th} sample from song A. $D(A,B)$ is a measure of the similarity between song A and song B.

This global timbre similarity measure is implemented in CUIDADO music browser (Pachet, La Burthe, Zils, and Aucouturier, 2004). The results for the query "Ahmad Jamal - L'instant de Verite" —a jazz piano recording returns similarity results which all contain *romantic-styled* piano. For example, New Orleans Jazz (*G. Mirabassi*), Classical Piano (*Schumann, Chopin*). It should be noted that some of the most interesting results are unexpected (different genres and cultural backgrounds).

Finding an evaluation metric for this type of system would be difficult. The MIR community has 'hotly' debated the subject of evaluation. At this time standard test databases need to be developed in order to compare different techniques. There is also the question of what exactly defines similarity? Comparing to hand segmented/clustered results might not be adequate since false-negatives might be generated for unexpected results.

References

Aucouturier, J.J., F. Pachet, and M. Sandler. 2005. The way it sounds: Timbre models for analysis and retrieval of polyphonic music signals. *IEEE Transactions of Multimedia* .

Grey, J.M. 1975. An exploration of musical timbre using computer-based techniques for analysis, synthesis and perceptual scaling. Ph.D. thesis, Dept. of Psychology, Stanford University.

Herrera-Boyer, P., G. Peeters, and S. Dubnov. 2003. Automatic Classification of Musical Instrument Sounds. *Journal of New Music Research* 32(1):3–21.

Pachet, F., A. La Burthe, A. Zils, and J.J. Aucouturier. 2004. Popular music access: The Sony music browser. *Journal of the American Society for Information Science and Technology* 55(12):1037–1044.

Wessel, D. 1979. Timbre space as a musical control structure. *Computer Music Journal* 3(2):45–52.

Figures

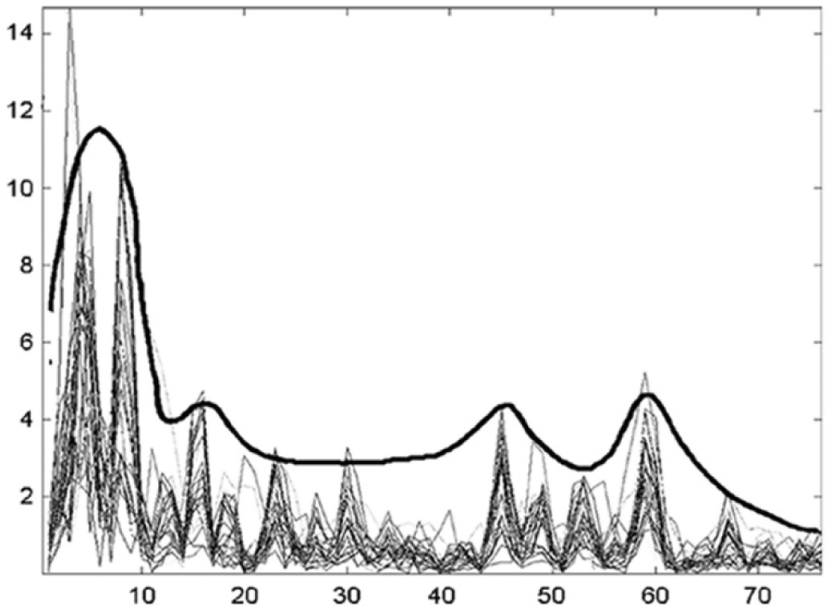

Figure 1: Global Spectral Shape(Aucouturier 2005)

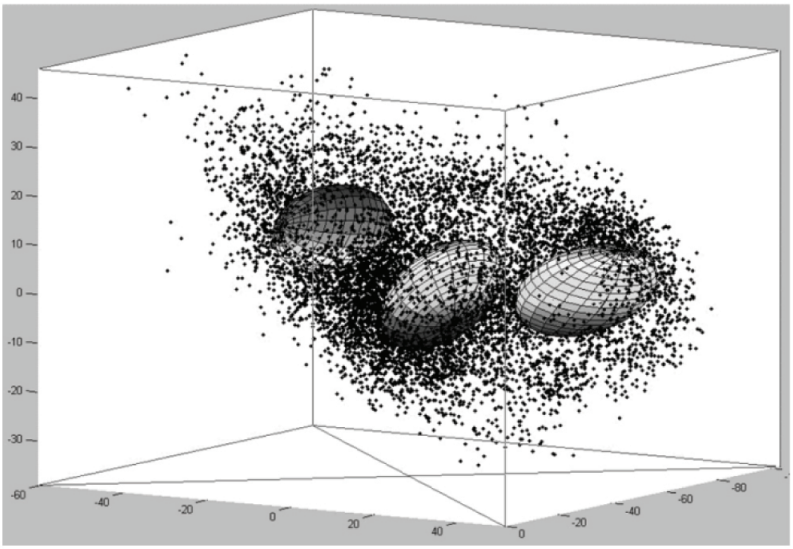

Figure 2: GMM Clustering (Aucouturier 2005)