Rebecca Fiebrink
10 March 2005

An Overview of Work by Pachet and Aucouturier on Timbre Similarity

Introduction

In the last few years, there has been much research into developing frameworks for electronic music distribution and browsing systems that can best meet users' needs. One vein of this research involves computing perceptually-based measures of music similarity. François Pachet and Jean-Julien Aucouturier, two researchers from the Sony Computer Science Laboratory in Paris, have contributed significantly to this body of work. This review outlines their main contributions to research on timbre similarity in particular.

Motivation

Aucouturier and Pachet's work on timbre similarity is motivated by the growing area of Electronic Music Distribution (EMD). EMD allows large numbers of music selections to be distributed to large user populations, but to be effective, it must be accompanied by means for users to efficiently browse through these electronic music collections. Aucouturier and Pachet argue that incorporating perceptual measures of musical similarity, specifically timbre similarity, can facilitate browsing through EMD systems. Musical taste is often associated with timbre, and timbre similarity is a useful and natural way to build relationships between music selections (Aucouturier and Pachet 2002a).

Initial Work

Their first work toward a timbre similarity system, published in 2002, involved the use of Mel Frequency Cepstral Coefficients (MFCCs) to compute a high-level timbre descriptor for each musical selection. This descriptor was built by finding the first eight MFCCs for every 50 milliseconds of audio and modeling the set of all MFCCs in the song as a 3-state Gaussian Mixture Model (GMM). The similarity between two songs is inversely related to the distance between the GMMs of the two songs, and this distance was computed by sampling using 1000 samples. The compact GMM information could be stored with each song in a database, and a similarity matrix could be computed offline for all pairs of songs for quick lookup by a user (Aucouturier and Pachet 2002a; 2002b).

Aucouturier and Pachet evaluated the performance of the 2002 system favorably. The system often judged selections by the same artist or in the same genre to have similar timbre, which is to be expected for any timbre similarity measure. Additionally, the system proposed timbre matches such as a piece by Beethoven with a Beatles song. The authors described such matches as "interesting," and argued that it was on those matches that would not be proposed based on artist or genre metadata alone that a timbre similarity measure proves most useful. There is difficulty inherent in testing a system such as this where no ground truth exists, but a study comparing this system to human listeners demonstrated agreement on similarity 80% of the time (Aucouturier and Pachet 2002a; 2002b).

CUIDADO and the Music Browser

This system was implemented within the European CUIDADO (Content-based Unified Interfaces and Descriptors for Audio/music Databases Available Online) project framework. This larger project, which lasted from 2001 to 2003, focused on music description structures, the development of extractors for high-level audio information, and the implementation of the Sound Palette and Music Browser applications (Vinet et al. 2002). 17, 075 popular music selections and associated metadata were accessible from the Music Browser, and work on the Browser continued through the duration of the CUIDADO project (Pachet et al. 2004). As part of their initial work, Aucouturier and Pachet (2002a; 2002b) integrated the timbre similarity measure so that users could directly search for songs that matched the timbre of a given song or generate playlists from a set of constraints, including constraints on timbre.

Similar Work Through 2003

Pampalk, Dixon, and Widmer's paper discussed five perceptual music similarity measures, one of which was Aucouturier and Pachet's 2002 system. While not all measures dealt specifically with timbre similarity, several also used MFCCs in similar computations. Aucouturier and Pachet's system appeared to work relatively slowly but performed better than average (Pampalk, Dixon, and Widmer 2003).

Improvements to the 2002 System

In 2004, Aucouturier and Pachet conducted a series of tests varying the algorithms and parameters used in their first system and evaluated the results within the Music Browser. They examined the extent of the improvements made possible by increasing the signal sampling rate and the GMM sampling rate, replacing GMM sampling with Earth Mover's Distance, adjusting the number of MFCCs and GMM components, and changing the analysis window size. Even when they used their discovered optimum parameter values, they found that the system performance improved by only 16%. They also experimented with using a variety of speech processing algorithms, which offered at best a 2% increase, and they tried using hidden Markov models instead of GMMs, which offered no improvement (Aucouturier and Pachet 2004a; 2004b).

As a result of these tests, they concluded that little improvement was likely past 65% precision using the existing system framework. It is notable that this precision value is somewhat conservative, as they only considered same-genre matches to be correct for the purposes of evaluation. However, false-positives still seemed to be an issue in the 2004 system, and the authors proposed that further improvements might best be made through a greater focus on timbre perception. For instance, some frames of an audio signal might be more important than others in determining a piece's perceived overall timbre quality, and some timbres may be more salient than others (Aucuturier and Pachet 2004a; 2004b).

Conclusions

The above studies suggest that timbre similarity systems, though in their infancy, can already provide us with effective and efficient supplements to existing music recommendation and browsing systems. However, existing systems based on MFCC

distributions are far from perfect, and their evaluation and improvement are frustrated by the difficulty of objective evaluation and by open questions regarding human timbre perception.

References

Aucouturier, J., and F. Pachet. 2002a. Finding songs that sound the same. *Proceedings of the First IEEE Benelux Workshop on Model Based Processing and Coding of Audio*, 91–98.

———. 2002b. Music similarity measures: What's the use? *Proceedings of the International Conference on Music Information Retrieval.*

———. 2004a. Improving timbre similarity: How high's the sky? *Journal of Negative Results in Speech and Audio Sciences* 1 (1).

———. 2004b. Tools and architecture for the evaluation of similarity measures: Case study of timbre similarity. *Proceedings of the International Conference on Music Information Retrieval*.

Pachet, F., A. La Burthe, A. Zils, and J. Aucouturier. 2004. Popular music access: The Sony music browser. *Journal of the American Society for Information Science* 55 (12): 1037–44.

Pampalk, E., S. Dixon, and G. Widmer. 2003. On the evaluation of perceptual similarity measures for music. *Proceedings of the 6th International Conference on Digital Audio Effects*, 6-12.

Vinet, H., P. Herrera, and F. Pachet. 2002. The CUIDADO project. *Proceedings of the International Conference on Music Information Retrieval*.