AMS and Genre Classification Using Timbre, Rhythm and Tempo Models

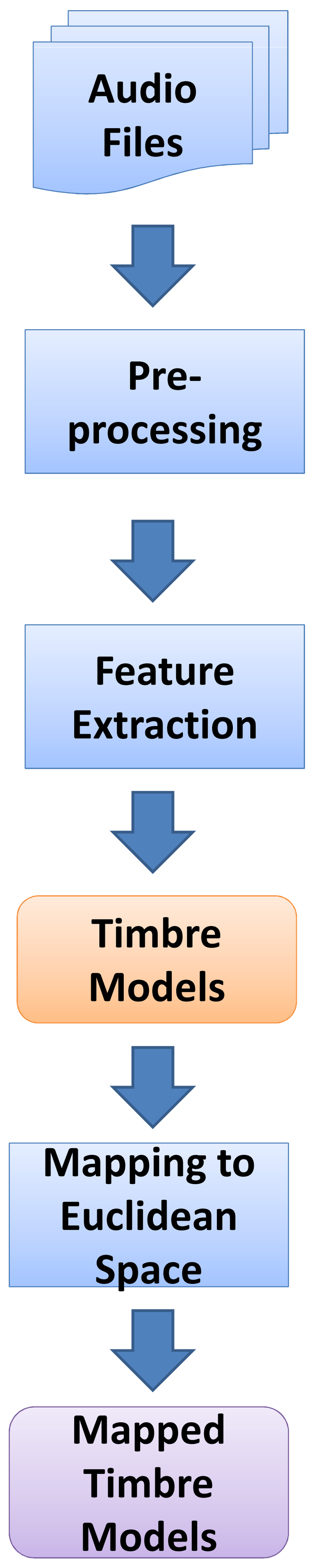
Franz de Leon, Kirk Martinez
Web and Internet Science Group • School of Electronics and Computer Science • University of Southampton
fadl1d09@ecs.soton.ac.uk

Summary

Audio features that approximate **timbre**, **rhythm** and **tempo** are used for genre classification and music similarity estimation. In our system, audio data are modelled as long-term accumulative distribution of frame-based spectral features. This is also known as the "bag-of-frames" (BOF) approach wherein audio data are treated as a global distribution of frame occurrences.

Feature Extraction

Timbre Component



•For each normalized track track, the audio signal is segmented into 23 msec. non-overlapping, Hanning-windowed frames.

•The Mel-frequency Cepstral Coefficients (MFCC) [1] and its time derivative (ΔMFCC) are computed for each song.

•The features are summarized as single multivariate Gaussian with full covaraince matrix.

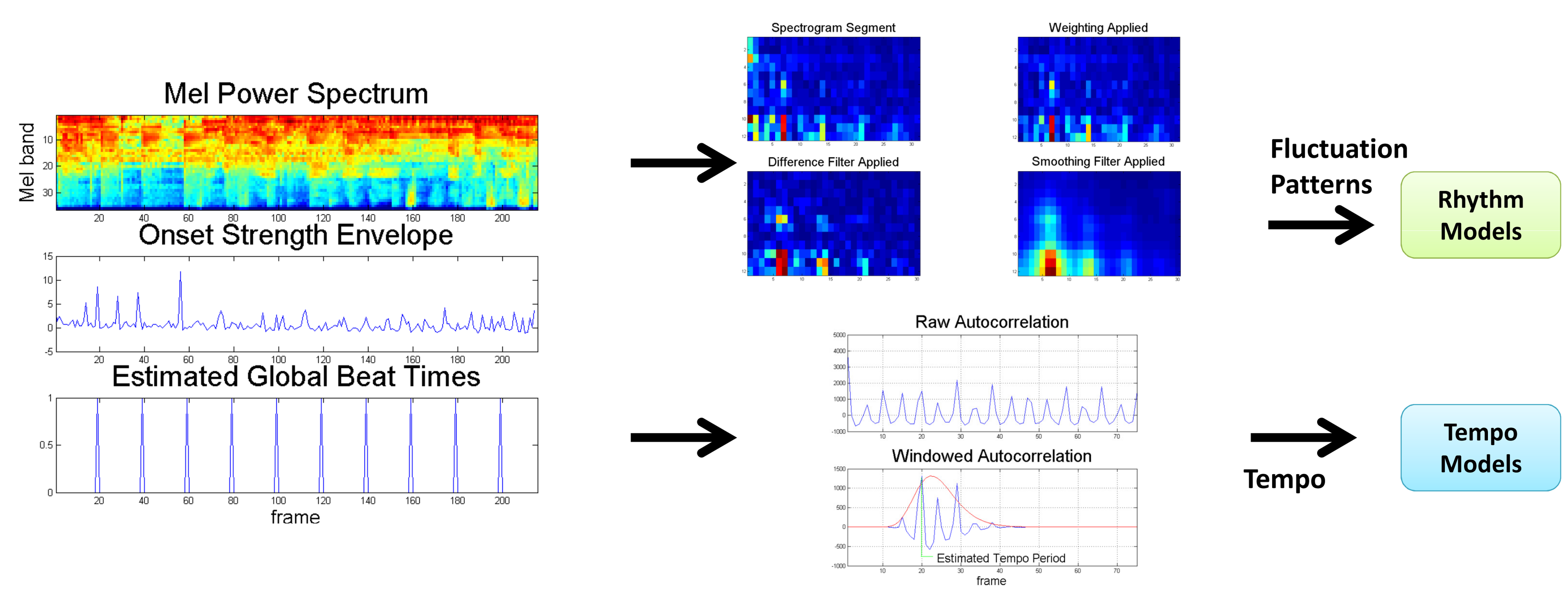
•The MFCC and ΔMFCC vectors are mapped to *k*-dimensional Euclidean space using a modified Fastmap [2] algorithm.

Rhythm Component

The rhythm component is based on the Fluctuation Patterns [3] (FPs) of the audio signal. Fluctuation patterns describe the amplitude modulation of the loudness per frequency band. For each frame, the fluctuation pattern is represented by a 12x30 matrix. The rows correspond to reduced Mel-frequency bin while the columns correspond to modulating frequency bands. To summarize the FPs, the median of the matrices is computed. Additional features derived are FP mean and FP standard deviation.

Tempo Component

The tempo component is derived from a technique using onset autocorrelation [4]. The tempo is computed by taking the first-order difference along time of a Mel-frequency spectrogram then summing across frequency. A high-pass filter is used to remove slowly-varying offsets. The global tempo is estimated by autocorrelating the onset strength and choosing the period with the highest windowed peak.

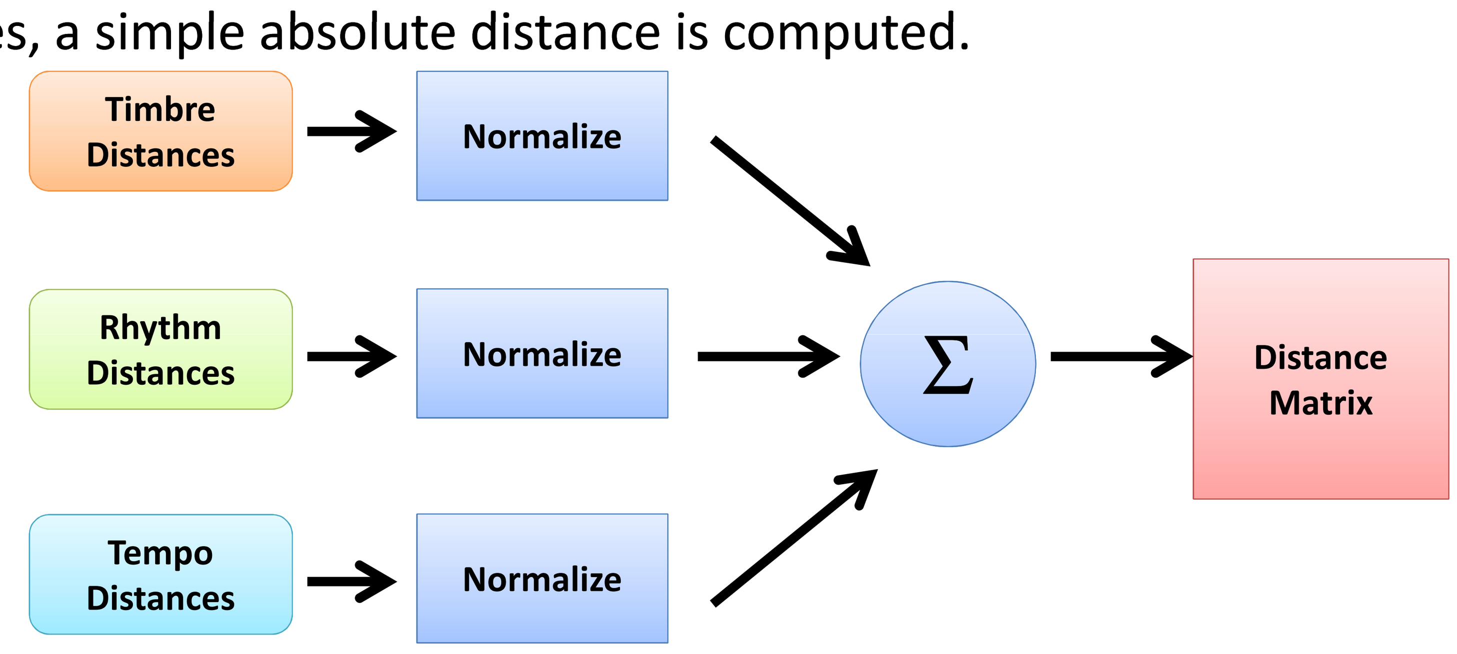


Methods

Audio Music Similarity Estimation

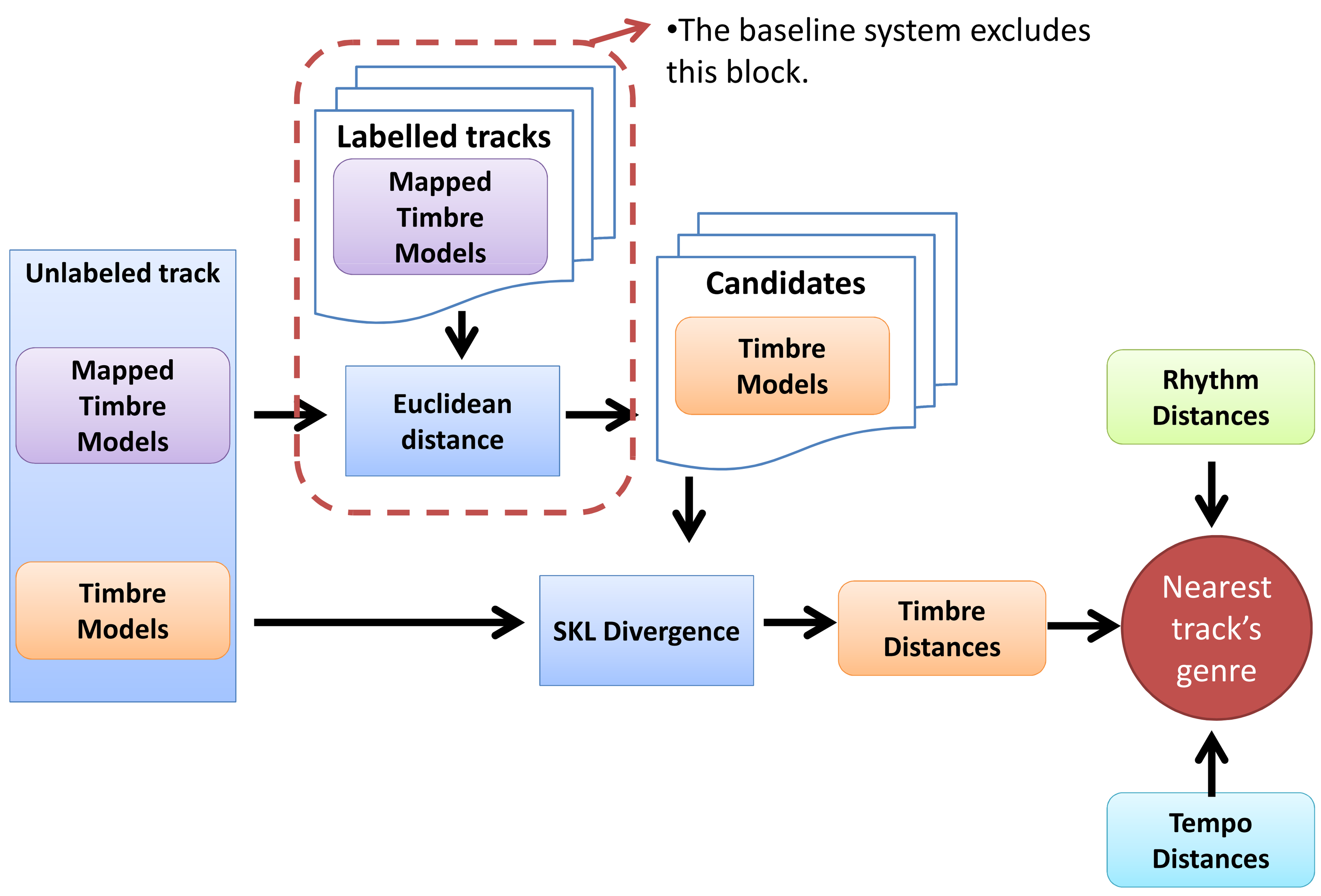
The timbre, rhythm, and tempo distances are calculated separately. Before they are combined, each distance component is normalized by removing the mean and dividing by the standard deviation of all the distances. The results are then weighted before being summed. Symmetry is obtained by summing up the distances in both directions for each pair of tracks [5].

Distances between timbres are computed by comparing the GMM models. We use symmetric Kullback-Leibler (KL) distance between two models [6][7]. The Euclidean distance is used to compute distance between rhythms. For tempo distances, a simple absolute distance is computed.



Genre Classification

Genre classification is done using nearest neighbours based on the weighted sum of timbre, rhythm and tempo distances.



Results

This submission is an updated version of the algorithm submitted to MIREX 2011 AMS task. From the MIREX 2011 data, the objective results are highly correlated with the human evaluation grades. Both objective and subjective results show that returning the 5 closest songs to a given query, 50% of the candidate songs belong to the same genre.

Acknowledgements
Franz de Leon is supported by the Engineering Research and Development for Technology Faculty Development Program of the University of the Philippines, and DOST

References

[1] B. Logan and A. Salomon, "A Music Similarity Function Based on Signal Analysis," in *IEEE International Conference on Multimedia and Expo*, pp. 745-748, 2001.
[2] D. Schnitzer, A. Flexer, and G. Widmer, "A Filter-and-Refine Method for Fast Similarity Search in Millions of Tracks," in *ISMIR 2009*, 2009, no. April, pp. 537-542.
[3] E. Pampalk, "Computational Models of Music Similarity and their Application in Music Information Retrieval," Vienna University of Technology, 2006.
[4] D. P. W. Ellis and G. E. Poliner, "Identifying `Cover Songs' with Chroma Features and Dynamic Programming Beat Tracking," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, 2007, p. IV-1429-IV-1432.
[5] E. Pampalk, "Audio-Based Music Similarity and Retrieval : Combining a Spectral Similarity Model with Information Extracted from Fluctuation Patterns," in *Submission to MIREX 2006*, 2006.
[6] W. Penny, "Kullback-Liebler Divergences of Normal, Gamma, Dirichlet and Wishart Densities," 2001 .
[7] C. Charbuillet, G. Peeters, S. Barton, and V. Gouet-Brunet, "A fast algorithm for music search by similarity in large databases based on modified Symetrized Kullback Leibler Divergence," in *2010 International Workshop on Content Based Multimedia Indexing (CBMI)*, 2010, pp. 1-6.