

Computation of Rhythmic Similarity in Music

Carmine Casciato, McGill University

Temporal patterns of sound are everywhere. The question of comparing them is a non-trivial one. Timbre and tempo differences must be considered. When these parameters are normalized, the next issue is what constitutes similarity between rhythms. The various approaches reviewed here all use alternate representations to categorize the similarity of different rhythms, .

Within the sphere of music information retrieval, there has been little explicit research in rhythmic similarity, although one can consider any work in rhythmic tracking or genre classification to be in the same sphere. The following works dealt with rhythmic similarity issues directly.

Jouni Paulus and Anssi Klapuri (2002) implement a system that focuses on recorded music that contains percussion sounds. The signal is initially re-synthesized using a sinusoid plus noise spectrum model to remove harmonic partials, thereby leaving typically noisy percussion sounds. The proposed rhythmic segmentation module works with unprocessed audio as well. Using a series of filter banks to decimate and downsample the signal into vectors of amplitude envelopes according to the respective passband, a quasi-autocorrelation is performed across all vectors to determine salient periodicities. Feature extraction is performed on pattern segments resulting in matrices of loudness, spectral centroid, and mel-frequency cepstral coefficients (MFCCs) for each segment. The matrices are then compared using Dynamic Time Warping (DTW) “by trying to find an optimal path through a matrix of points representing all the possibly time alignments between the feature” matrices. DTW is used instead of Hidden Markov Models because the comparison is between individual data sets rather than predetermined rhythm classes. The authors proceed to explain several tests performed on various data sets and different aspects of the model. It is difficult to ascertain the appropriateness of the model as the entire implementation is not thoroughly tested. One notable finding is the robustness of the normalized spectral centroid feature in rhythmic similarity comparisons.

Jonathan Foote et al. (2002) use a spectral representation to compare audio examples. A Fast Fourier Transform of the audio signal results in vectors of spectral information that are embedded in a ‘similarity matrix’ which shows the spectral similarity between different instants in one single audio segment. The ‘beat spectrum’ value is then obtained from the matrix by summing along diagonals. The authors contend this gives a reasonable tempo categorization of audio. The two experiments outlined include a measure of different distance metrics between beat spectrum vectors: Euclidean distance, cosine angle of vectors, and the Fourier beat spectral coefficients. The latter two metrics perform over 90% for categorizing 15 audio excerpts of 4 songs into 5 relevance

sets. The authors end the paper by proposing supervised learning techniques for improving the algorithm and also proposing possible applications for the technique, including automatic sequencing of rhythmically similar music using beat spectrums.

Daniel Ellis and John Arroyo (2004) present a different approach to rhythmic similarity using Principle Component Analysis (PCA). They reduce the drum beats of 100 popular songs in MIDI format to 3 MIDI drum sounds and apply a PCA analysis, which reduces high-dimensional data to weighted sums of basis vectors. Nonetheless, there are only two dimensions in their representation, namely time and drum note. The segmentation of rhythms is done through the autocorrelation of inter-onset intervals. Four tempo hypotheses are culled from this step corresponding to the four highest peaks of the autocorrelation data. In order to align rhythm patterns for comparison, they average the test input itself to arrive at a reference pattern, and then compare the input to this reference pattern. Despite this priming, the system performs poorly in categorizing 100 songs into 10 genres, correctly categorizing only 20% of the songs. This is partly due to the usage of 2s segments of music which was not enough to encode and entire rhythm patterns in some examples. Nonetheless, the authors feel that the “eigenrhythm” framework does encode the general ‘feel’ of the rhythms and suggest possible future improvements.

Departing from the field of music-technology research, we arrive at the the work of Godfried Toussaint, a member of the Faculty of Computer Science here at McGill. Toussaint’s interests in rhythm and computational geometry combine in his analysis of the similarity of certain clave rhythms that are ubiquitous in African, Brazilian, and Cuban music (2002). By representing these rhythms in several different visual representations, including convex polygons and box notation, certain latent characteristics come to the surface. Using the latter representation, Toussaint evaluates several similarity distance measures, including Hanning Distance, Euclidean distance, and interval vector distance (which is essentially inter-onset interval times vectorized), and creates minimum spanning trees from these distances to visualize rhythm similarity and ancestry. The work is intriguing in its implications, but presently lacks an implementation using audio signals.

For work on rhythmic similarity in music perception and musicology, the reader is referred to Gabrielsson (1973), Powel and Essens (1985), and Lerdahl and Jackendoff (1983).

References

Ellis, D., and J. Arroyo. 2004. Eigenrhythms: Drum pattern basis sets for classification and generation. In *Proceedings of the International Conference on Music Information Retrieval*.

Foote, J., M. Cooper, and U. Nam. 2002. Audio retrieval by rhythmic similarity. In *Proceedings of the International Conference on Music Information Retrieval*.

Gabrielsson, A. 1973. Similarity ratings and dimension analyses of auditory rhythm patterns. *Scandinavian Journal of Psychology* 14: 138–60.

Lerdahl F., and R. Jackendoff. 1983. *A generative theory of tonal music*. Cambridge: The MIT Press.

Paulus, J., and A. Klapuri. 2002. Measuring the similarity of rhythmic patterns. In *Proceedings of the International Conference on Music Information Retrieval*.

Powel, D., and P. Essens. 1985. Perception of temporal patterns. *Music Perception* 2(4): 411–40.

Toussaint, G. 2002. A mathematical analysis of African, Brazilian, and Cuban clave rhythms. In *Proceedings of BRIDGES: Mathematical Connections in Art, Music, and Science*.