Rhythmic Similarity through Elaboration

Mitchell Parry and Irfan Essa
College of Computing / GVU Center
Georgia Institute of Technology
Atlanta, GA 30332-0280
{parry,irfan}@cc.gatech.edu

Abstract

Rhythmic similarity techniques for audio tend to evaluate how close to identical two rhythms are. This paper proposes a similarity metric based on rhythmic elaboration that matches rhythms that share the same beats regardless of tempo or identicalness. Elaborations can help an application decide where to transition between songs. Potential applications include automatically generating a non-stop music mix or sonically browsing a music library.

1 Introduction

This paper proposes a similarity metric based on rhythmic elaboration that matches rhythms that share the same beats regardless of tempo or identicalness. Using this metric, we can identify relatively complex or elaborated rhythms and repeated rhythmic patterns within a song or collection. Rhythmic similarity metrics can be applied at multiple levels. Some techniques (Tzanetakis and Cook, 2002; Foote et. al., 2002) derive a single representation for an entire song. These techniques aid in song retrieval and automatically ordering songs in a play list. Other techniques, such as Paulus and Klapuri (2002), compare rhythms at the measure level using low-level features. Their technique allows approximate matches and small tempo changes. Foote and Cooper (2001) reveal high level repetition in music by visualizing a self-similarity matrix. Tanguiane (1993) has proposed symbolic music to compare rhythms with a different number of beats and equal total duration. He uses the concept of elaboration to label rhythmic phrases. An elaboration of a rhythm is a rhythm that contains all the same onsets. He expresses rhythms as bit vectors, $R_k = < r_1, r_2, ..., r_n >$, and identifies elaborations by the simple relation: R_i is an elaboration of R_j if $R_i \cdot R_j = R_j \cdot R_j$. Clearly, if R_i and R_j are elaborations of each other, they are identical. Elaborations applied to audio help decide where to transition between songs. Potential applications include automatically generating a non-stop music mix or sonically browsing a music library.

2 Methods

We use a variant on Scheirer's (1998) beat analysis that uses the discrete wavelet transform as proposed by Tzanetakis et. al. (2001) instead of a filterbank. Starting with a 22 kHz mono signal we output a 172 Hz beat envelope. We chose to partition the signal into measure-length segments. Paulus and Klapuri (2002) describe a probabilistic algorithm for determining the measure length from an audio signal. Using this measure length, we divide the song into non-overlapping adjacent segments.

2.1 Similarity

We adapt Tanguiane's (1993) method for finding elaborations to our amplitude envelope segments. To account for phase and amplitude differences, each pair of segments are shifted according to the peak in the cross-correlation and normalized by their maximum value. Then all pairs are evaluated by the asymmetric relation,

$$elab(s_i, s_j) = 1 - \frac{\min(s_i \cdot s_j, s_j \cdot s_j)}{\max(s_i \cdot s_j, s_j \cdot s_j)}. \quad (1)$$

Segment s_i is an elaboration of segment s_j if $elab(s_i, s_j)$ is near zero. If s_i and s_j are identical, $elab(s_i, s_j)$ and $elab(s_j, s_i)$ is zero. For comparison, we use the symmetric cosine distance,

$$dcos(s_i, s_j) = 1 - \frac{s_i \cdot s_j}{\|si\| \|sj\|}. \quad (2)$$

The elaboration and cosine relation generates a matrix that represents the entire song.

2.2 Complexity

Complexity can be evaluated on two levels: within a song and between songs. To evaluate the complexity of a rhythmic segment relative to the rest of the song, we can use the row and column sums in our elaboration matrix. The difference of these two sums is positive for segments that tend to have elaborations and negative for segments that tend to be elaborations. We use the function,

$$complexity(s_i) = \sum_j elab(s_i, s_j) - elab(s_j, s_i). \quad (3)$$

Complexity can also be evaluated between songs (i.e. the complexity in transitioning between two segments). The application here is to find good transitions between songs. The best candidates for transitions are those that are elaborations of each other.

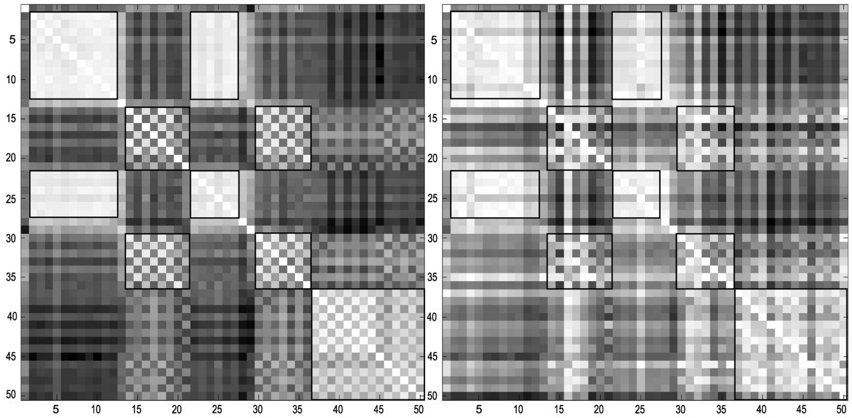

Figure 1: Cosine (left) and elaboration (right) similarity matrices

For this application, we do not care about the direction of the elaboration. We assume that it is equally pleasing to transition to a more elaborated rhythm or a simpler rhythm. Therefore we use a transition rating where ratings near zero are ideal:

$$trans(s_i, s_j) = \min(elab(s_i, s_j), elab(s_j, s_i)). \qquad (4)$$

3 Results

By visualizing the elaboration matrix within a song we can show the added information it provides. Figure 1 shows the cosine similarity matrix (left) and the elaboration matrix (right) for the first half of "Down In It (Demo)" by Nine Inch Nails. Time runs down and to the right. White indicates high similarity (value near zero). Repetition can be seen on both matrices where white rectangles appear off the diagonal, for instance, segments 2-12 and 22-27. Another more subtle repetition occurs in the checkerboard pattern during segments 14-21 and 30-36. These regions are outlined in black for comparison between the matrices. The cosine similarity matrix is generally dark where repetition does not occur. The elaboration matrix contains white rows and dark columns that indicate a relatively elaborated pattern, and vice versa for relatively simple patterns. For instance, row 35 is bright and column 35 is dark. Instances where $elab(s_i, s_j)$ and $elab(s_j, s_i)$ is relatively dark, indicate that these rhythms are wholly different (i.e. they each contain beats that the other does not).

In order to compare rhythms between songs, we use a collection of 38 songs from the Dave Matthews Band. We extract the first measure from each song and time stretch using commercial software so that all rhythms have the same length. Then we generate an elaboration matrix that compares every song to every other song. Figure 2 shows an example of a good transition, $trans = 0.01$, (left) and a bad transition, $trans = 0.53$, (right). For the good example, notice that all of the beats in the lower rhythm are contained in the top rhythm. Therefore, blending these rhythms will provide a smooth transition. In the bad example, notice that both rhythms contain beats that do not exist in the other. For instance, the top rhythm has a beat at 25 and the bottom has a beat at 330.

4 Conclusions

This paper proposes elaboration as a rhythmic similarity metric. We show that it provides more information than the cosine distance metric. Rhythmic elaboration can be used to identify rhythms that share the same beats as a target rhythm. Elaborations could aid applications that automatically transition between multiple audio sources, such as a non-stop music mix or sonic browser for music libraries.

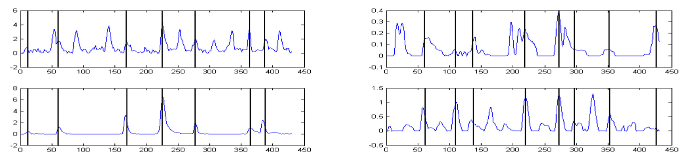

Figure 2: A good transition, $trans = 0.01$, (left) and a bad transition, $trans$ = 0.53, (right)

References

Foote, J. & Cooper, M. (2001). Visualizing Musical Structure and Rhythm via Self-Similarity. In Schloss, A., Dannenberg, R., & Diessen, P. (Eds.), *Proceedings of the International Computer Music Conference*, 27, (pp. 419–422). San Francisco, CA: International Computer Music Association.

Foote, J., Cooper, M., & Nam, U. (2002). Audio Retrieval by Rhythmic Similarity. In Fingerhut, M. (Ed.), *Proceedings of the International Conference on Music Information Retrieval*, 3, (pp. 265–266). Paris: IRCAM – Centre Pompidou.

Paulus, J. & Klapuri, A. (2002). Measuring the Similarity of Rhythmic Patterns. In Fingerhut, M. (Ed.), *Proceedings of the International Conference on Music Information Retrieval*, 3, (pp. 150–156). Paris: IRCAM – Centre Pompidou.

Scheirer, E. (1998). Tempo and Beat Analysis of Acoustic Musical Signals. *Journal of the Acoustical Society of America*, 103(1), 588–601.

Tanguiane, A. (1993). *Artificial Perception and Music Recognition*. New York: Springer-Verlag.

Tzanetakis, G. & Cook, P. (2002). Musical Genre Classification of Audio Signals. *IEEE Transactions on Speech and Audio Processing*, 10(5), 293–302.

Tzanetakis, G., Essl, G., & Cook, P. (2001). Audio Analysis using the Discrete Wavelet Transform. In *Proceedings of WSES International Conference on Acoustics and Music: Theory and Applications*.