Rhythmic Similarity
-- a quick paper review

Presented by: Shi Yong

March 15, 2007

Music Technology, McGill University

Contents

Introduction

- Music can be looked from different aspects:
  - Melody
  - Harmony
  - **Rhythm**
  - Instrumentation
  - Form
  - Etc..
- Rhythms similar / dissimilar
  - Very easy for human - perceptually
  - Not so easy for computer - quantitative measuring

Introduction

- If the rhythmic similarity can be quantitatively measured by computer, so what's the usefulness?
  - Automatic ranking in huge music collection
  - Musical database searching
  - Music context analysis
  - Musical genre classification
  - Etc.

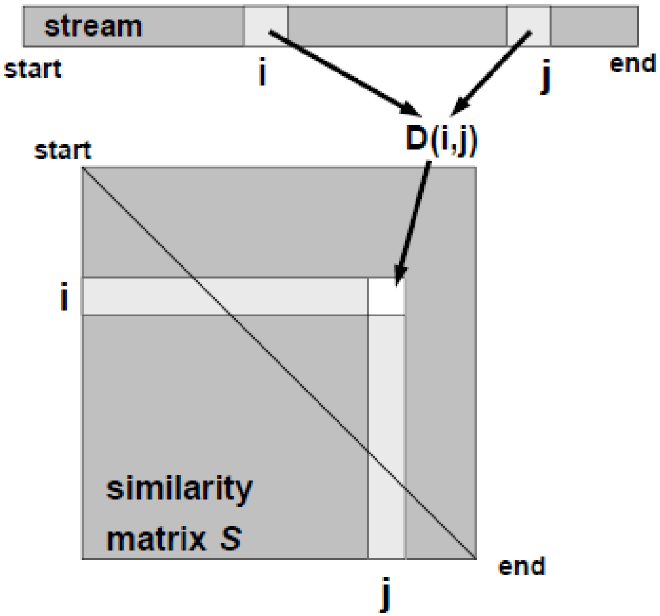
Example I – Foote's work (2001,2002)

- Key points
  - Novel approach to characterize the rhythm and tempo of music
    - Beat Spectrum
    - Beat Spectrogram
  - Measure the rhythmic similarity by distance of two beat spectra

Calculate Beat Spectrum

- Extract feature vectors from the audio stream
  - 256 samples frame wide
  - 50% overlapping
  - FFT and Power spectrum
- Cosine distances of all pairwise combinations of feature vectors

$$D_C(i,j) \equiv \frac{\mathbf{v}_i \bullet \mathbf{v}_j}{\|\mathbf{v}_i\| \|\mathbf{v}_j\|}$$



Similarity matrix calculation

Similarity Matrix

- A matrix S is constructed by all distance values in a signal
- Visualization: whiter regions = higher similarity

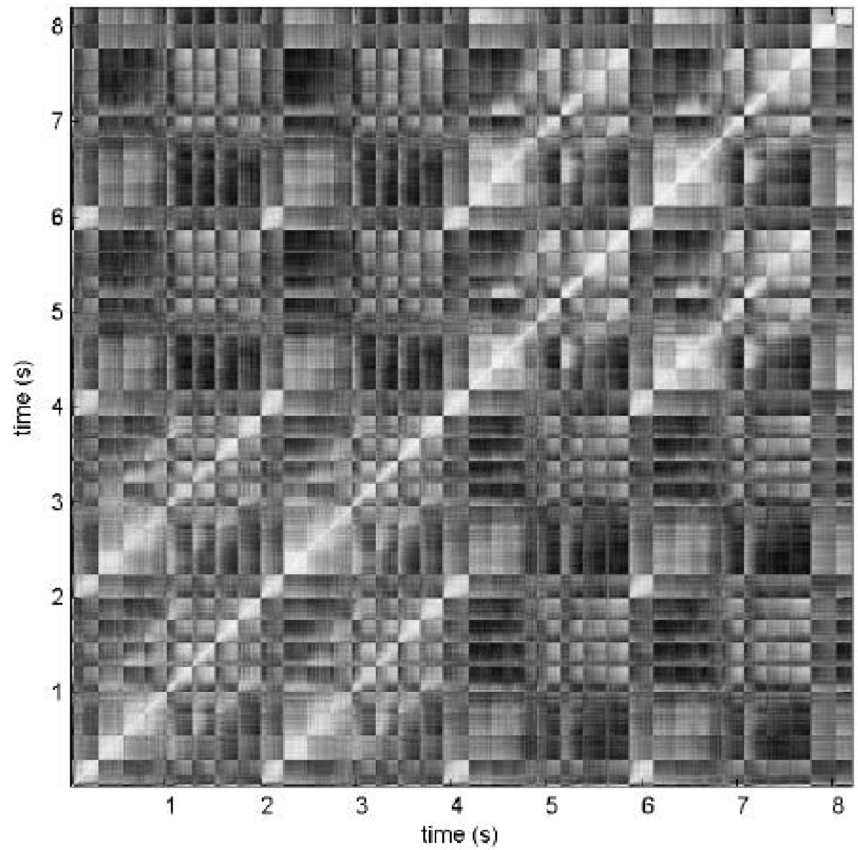


Figure 1. Similarity matrices for Bach's *Prelude No. 1 in C Major*, BVW 846. Performance: Glenn Gould.

Deriving the Beat Spectrum

- Beat Spectrum B(l) is a measure of self-similarity as a function of the time lag
- A simple estimation: summing S along the diagonal:

$$B(l) \approx \sum_{k \subset R} \mathbf{S}(k, k+l)$$

- More robust estimation comes from the autocorrelation of S
- Beat Spectrogram = Beat Spectrum over successive windows

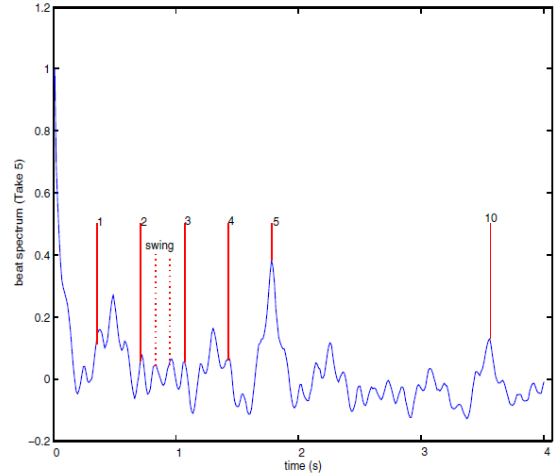


Figure 3. Beat spectrum of jazz composition *Take 5*

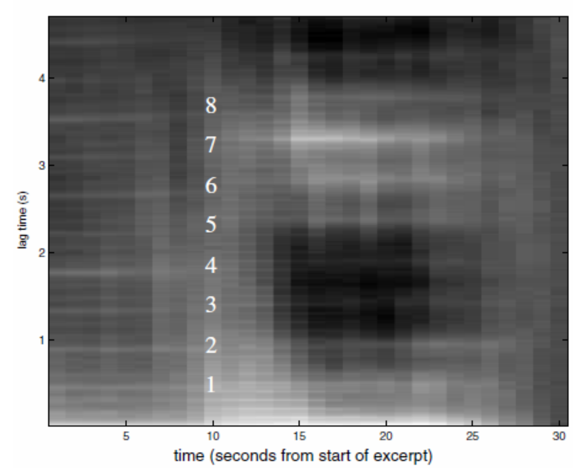


Figure 4. Beat spectrogram of Pink Floyd's *Money* (excerpt), showing transition from 4/4 to 7/4 time

Foote 2001, 2002

Measuring Rhythmic Similarity

- For two pieces, we have two beat spectra $B_1(l)$ and $B_2(l)$, where l is lag time (discrete and finite).
- The Rhythmic Similarity can be measured by the distance of two L-dimensional vectors
  - Squared Euclidean Distance
  - Cosine Distance
  - Cosine Distance of Fourier Beat Spectral Coefficients
  - Others
- Experiments were designed to evaluate the performance of different distance functions.

Foote 2001, 2002

Experiments

- In one experiment, it shows the Euclidean distance is also a measure of tempo difference.
- In another experiment, it shows the Cosine distance outperforms the squared Euclidean distance

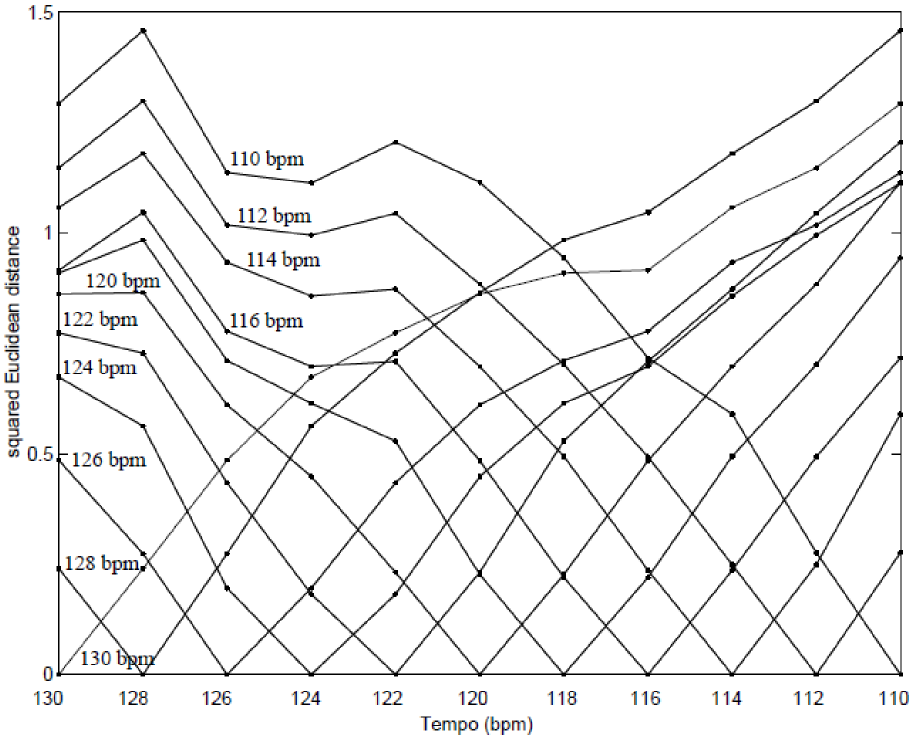
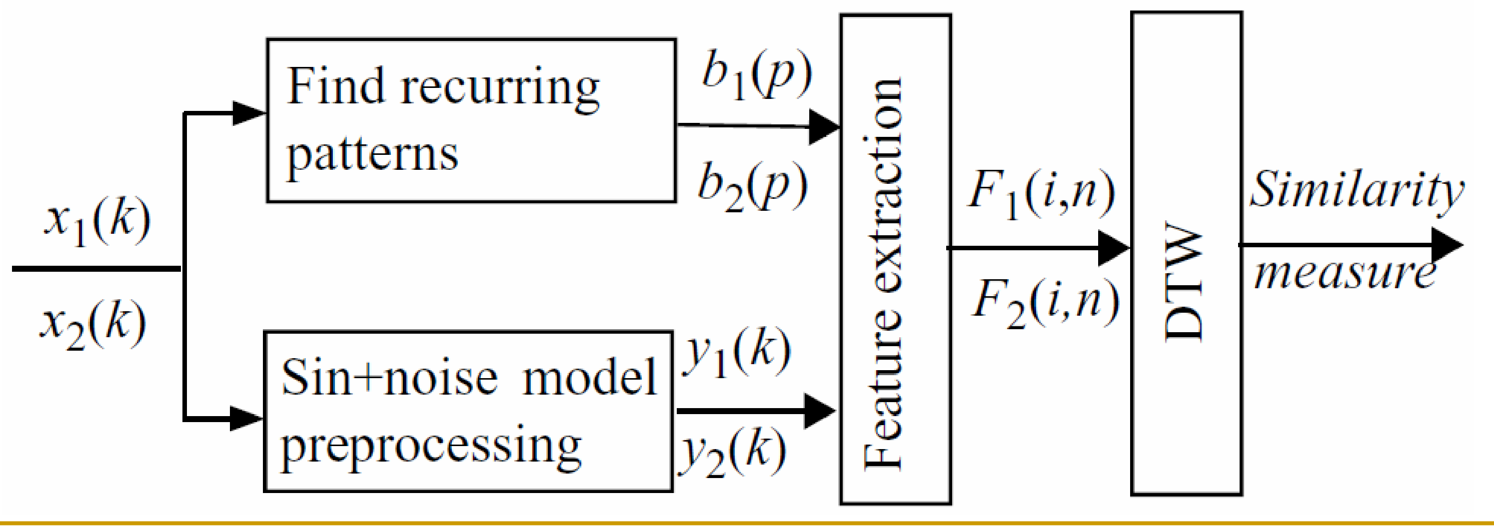


Figure 5. Euclidean Distance vs. Tempo

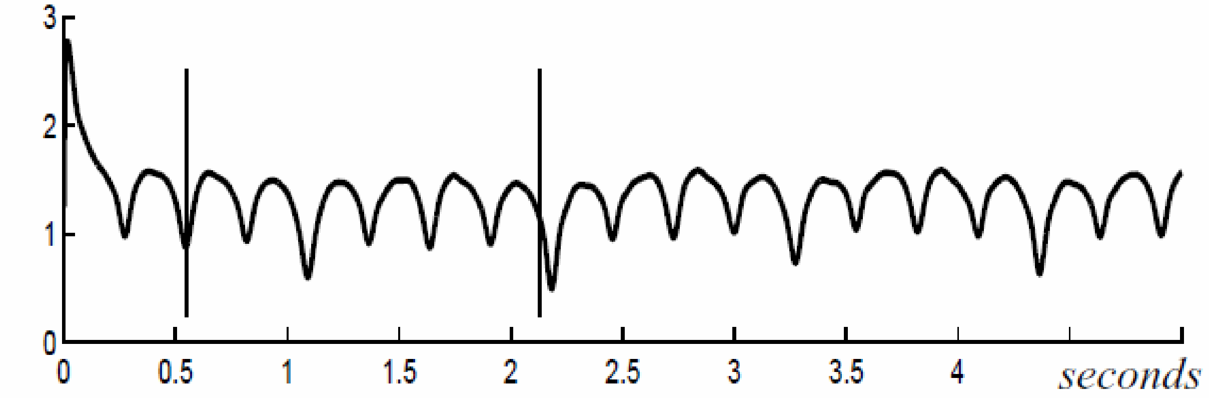
Example II – Paulus's work (2002)

- A system that measures the similarity of two arbitrary rhythmic patterns
  - Preprocessing (optional)
  - Rhythmic pattern segmentation
  - Features extraction
  - Similarity measuring



Pattern Segmenting

- The amplitude envelop is obtained from the audio stream by a set of processing methods
  - Normalizing, filter bank, half-wave rectify, square, decimation, low-pass, dynamic compression
- A periodicity analysis algorithm is then performed on the envelop signals to calculate the intermediary signal $s(\tau)$, which is used for musical meter estimation.



Pattern Segmenting

- Musical meters are estimated at three levels:
  - Tatum – the shortest duration
  - Tactus – beat
  - Musical measure
- Tatum period:
  - S(f) is calculate as the DFT of $s(\tau)$
  - "Tatum period is the inverse of the frequency corresponding to the maximum value" of $\sqrt{f} \times S(f)$
- Tactus period and musical measure period are estimated from $s(\tau)$ based on three probability distributions.
- A list of pattern boundaries are then produced, and one pattern can be isolated for further feature extraction

Feature Extraction

- Three features are extracted from one pattern – which is a series of overlapped frame.
  - Loudness – mean square energy of one pattern
  - Brightness – spectral centroid (using a logarithmic frequency scale)
  - MFCCs – 15 coefficients
- To avoid the absolute “tone color”, all features are normalized so that only the “up/down deviations” are remained
- Normalized feature matrix

$$F(i,n) = \begin{bmatrix} L'_1 & SC'_1 & \boldsymbol{MFCC'_1} \\ \dots & \dots & \dots \\ L'_n & SC'_n & \boldsymbol{MFCC'_n} \end{bmatrix}$$

Similarity Measuring

- Feature vector sets of two rhythmic patterns, F1(i,n) and F2(i,n), are matched by Dynamic Time Warping (DTW) algorithm
  - *“Dynamic time warping is an algorithm for measuring similarity between two sequences which may vary in time or speed.” – wikipedia*
- The similarity measure is given by

$$S(\boldsymbol{F}_1, \boldsymbol{F}_2) = \frac{\sqrt{N^2 + M^2}}{C(N, M)}$$

Results

- Pattern Segmenting
  - Tactus periods: 67% correct rate
  - Musical measure length: 77% correct rate
- Similarity Measuring
  - High similarity is assigned to the same rhythms performed with different drum sets

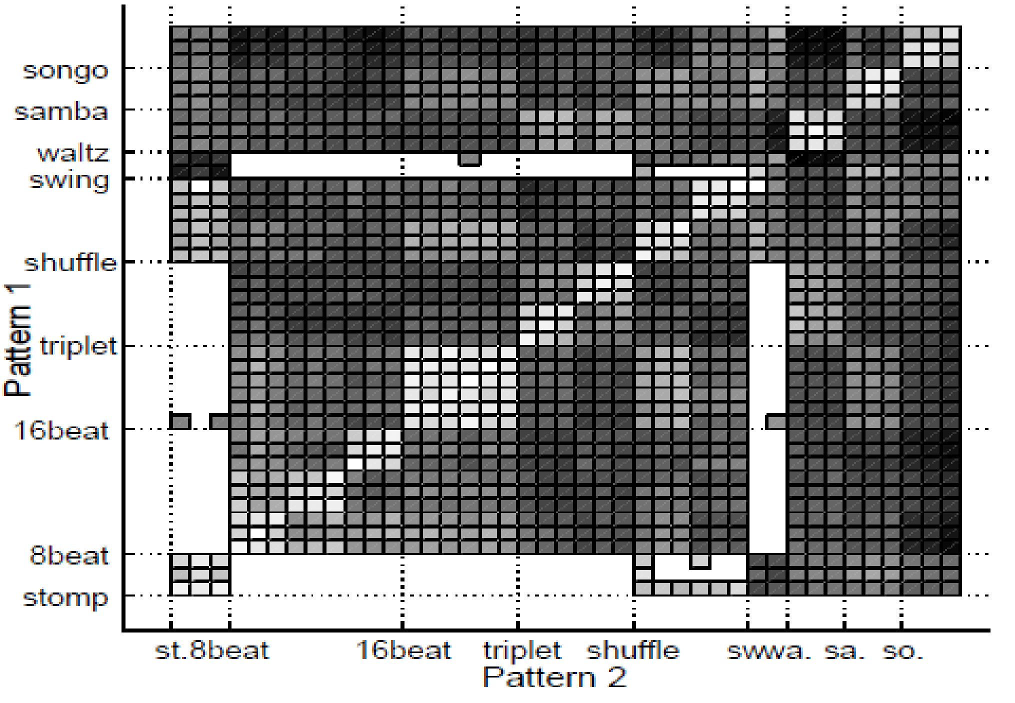


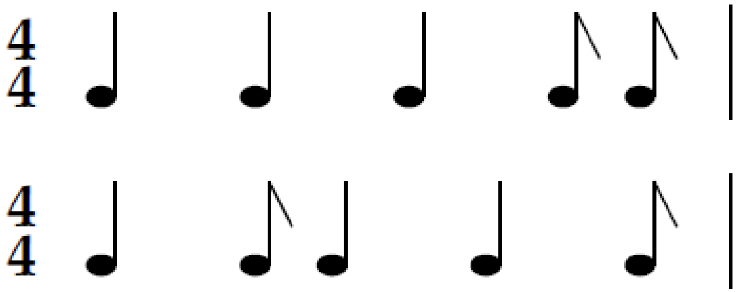
Figure 5. Calculated similarity measures for drum patterns.

14 rhythmic patterns performed by three different sound sets

Example III Dixon's work

- Key points:
  - A new way to characterize music by "typical bar-length rhythmic patterns"
  - Using it in music genre classification (ballroom dance music for this paper)

Temporal Sequence



Cha Cha – above

Rumba – below

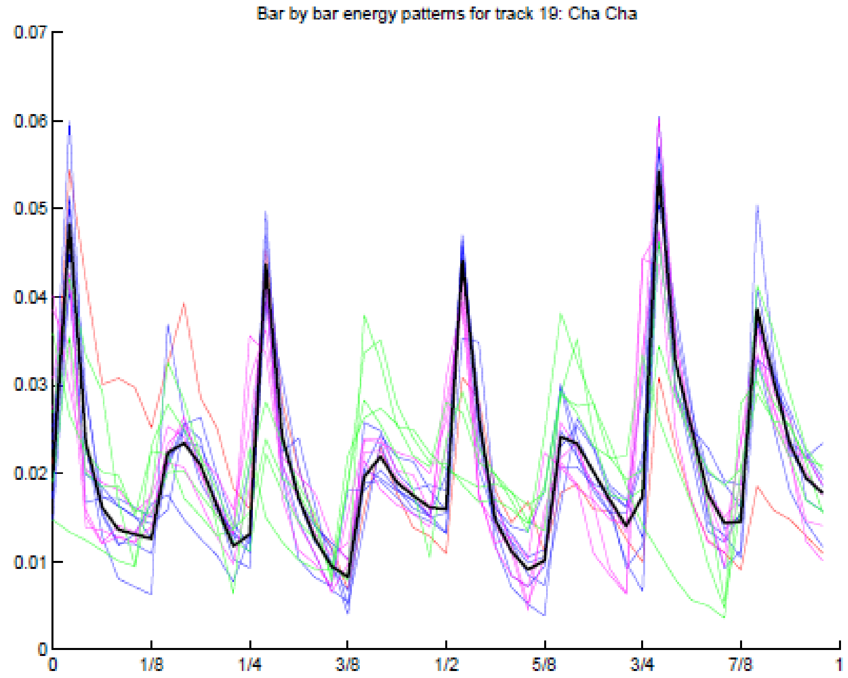
- The different genres of ballroom dance music are distinguished by the temporal sequence
- For genre classification purpose, the task is to automatically extract the rhythmic patterns from audio signal and compare the similarities

Main Steps

- First, the amplitude envelopes are extracted from a number of bar-length patterns
- Then, by using k-means clustering (k = 4), the most prominent rhythmic pattern is found by the largest cluster
- For similarity measuring – the distance of two patterns can be calculated by Euclidean distance
- For genre classification – the rhythmic pattern of each piece is used as a feature vector

Pattern Examples

- The amplitude envelope of fifteen bars of a Cha Cha excerpt
- Color curves are clusters belong to each bar
- Thick black curve is the largest cluster, defined as the typical pattern



Genre Classification

- Rhythmic pattern is used, alone or in conjunction with other feature set, for genre classification (dance music)
  - Rhythmic pattern
  - Features derived from rhythmic patterns:
    - Mean amplitude of the pattern
    - Maximum amplitude of the pattern
    - Standard deviation of the pattern
    - Etc.
  - Other automatically calculated feature set:
    - Features derived from the periodicity histogram
    - Features derived from the inter-onset interval histograms
    - Etc.
  - Measured tempo
- Classification rate
  - 50% - rhythmic pattern used alone (baseline is 16%)
  - 84% - when other automatically calculated features are included
  - 96% - when measured tempo is included

Conclusion

- Normal distance functions are used in ex. 1 and ex. 3, while in ex. 2, Paulus uses DTW to handle patterns with different lengths.
- Features extracted from both frequency domain (ex. 1 & ex. 2) and time domain (ex. 3) have been successfully tested
- Pattern segmentation is not easy (not mentioned in ex.1, but mentioned in ex.2 & 3)
- Tempo can be important for genre classification

References

- [Foote 2001] Foote, J. and S. Uchihashi. 2001. The Beat Spectrum: A New Approach to Rhythm Analysis. Proceedings of the International Conference on Multimedia and Expo .
- [Foote 2002] Foote, J., M. Cooper and U. Nam. 2002. Audio Retrieval by Rhythmic Similarity. Proceedings of the 3rd International Symposium on Musical Information Retrieval.
- [Paulus 2002] Paulus, J., and A. Klapuri. 2002. Measuring the Similarity of Rhythmic Patterns. Proceedings of the 3rd International Symposium on Musical Information Retrieval.
- [Dixon 2004] S. Dixon, F. Gouyon, and G.Widmer. Towards characterisation of music via rhythmic patterns. In ISMIR,Barcelona, Spain, 2004