VISUALIZATION OF METRE AND OTHER RHYTHM FEATURES

Enric Guaus, Eloi Batlle

Audiovisual Institute
Universitat Pompeu Fabra
eguaus,eloi@iua.upf.es

ABSTRACT

In this paper, we present a new rhythm, metre and BPM visualization tool. It is based on the proposed *Rhythm Transformation* that transforms audio data from time domain to a so called *rhythm domain*. The goal of this method is that data in *rhythm domain* can be interpreted as frequency domain information (for BPM detection) as well as time domain information (for metre detection). Some musical information, such as the metre (simple or compound, duple or triple, swinged or non-swinged), can be extracted from input audio. The method is based on the periodogram computation of the processed input data, and the different musical features are extracted by using well known techniques.

1. INTRODUCTION

Rhythm, melody and harmony are the most important topics in a musical composition or listening process. All of them are quite complex and their behaviour is still nor clear. Let us focus in the rhythm cognition process. In general terms, it is divided in three steps [1].

The first step in a rhythm recognition process deals about the *beat* and *tempo* detection. The *Beat* is the shortest periodic energy accent in a composition. As *beat* does not give us much musical information, some higher level structures, like *Tempo*, are needed. *Tempo* describes the speed of the beats, and it is usually a fixed property for the whole piece. *Tempo* can be interpreted as a fixed number (the BPM) or as a "how to play"(*allegro, andante*).

The second step in the rhythm recognition process is the *metre* detection, that is the organization of the strong and weak beats in a bar. According to the standards, the *metre* can be duple or triple, simple or compound. But whatever the *tempo* or *meter* are, we have not enough information for a successful recognition.

The third (and last) step in the process deals about *feeling*. A score usually shows a rigid version of the piece, and the musician should be able to change (perform) this rigid structure into a human sound, a sound transmitting feeling.

In a real situation, the listener perceives all these characteristics together. With all these three ingredients, the listener gets the whole idea about the *rhythm* of the piece. Perception is a very simple task for humans, but not the same for computers. This paper presents a metre and rhythm visualization method that could be usefull for further cognition research. Our algorithm works specially at the first two levels described above, but some incursions at the third level are shown too.

2. RELATED WORK

A lot of studies in rhythm features extraction can be found. At the first level in the rhythm recognition process, a successful approach is made by Eric D. Scheirer [2]: The system described in this paper worked as a beat and tempo extractor for any arbitrary polyphonic music. A sub-band frequency decomposition and the correlated energy of each band are used for the beat-tracking algorithm. Other successful works are made by Masataka Goto and Yoichi Muraoka [3]. They use onset-detection and linear-prediction techniques in their beat-tracking system, designed for general multimedia audio data.

At the second level in the rhythm detection process, the *Beat Spectrum* method was introduced by Foote and Uchihashi in [4]: The *Beat Spectrum* is a method for automatic characterization of the rhythm and the tempo of audio data. It is implemented as a measure of acoustic self-similarity as a function of time lag. The *Beat Spectrogram* is another tool that shows us the rhythm variations over time. Some distance measures and 2-D graphical representations are used for *Beat Spectrum* calculations. In a similar way, G. Tzanetakis et al presented the *Beat Histogram* in [5]. The *beat histogram* is a representation of the strongest periodicities from the processed input signal. These periodicities show such information like beat and tempo, so that some higher level rhythm information can be extracted.

At the top level of the rhythm perception process, some works have been done[6, 7]. George Tzanetakis shows us how to combine timbre textures, rhythmic content and pitch content in [8]. All these studies try to find a meeting point for musical knowledge and mathematical algorithms, a meeting point that sometimes showed to be inexistent. Finally, note that some of these works are subjected to some kind of restrictions. Our work is focused on realtime implementation with no signal input restrictions.

3. SYSTEM'S DESCRIPTION

Our work is clearly divided in three parts: the *Rhythm transformation* computation, the *Rhythm domain* visualization and the *Beatedness* calculation as an application of the global process.

3.1. Rhythm transformation

Most of the beat tracking systems compute the frequency analysis of the input signal and search for the common energy periodicities through different (linear o mel-frequency based) sub-bands. The energy's periodicity search is usually implemented as a bank of resonators and represented as a Beat Spectrum or as a Beat Histogram. Our system is slightly different: the periodogram is calculated for each sub-band and, finally, a weighted sum is implemented for a global rhythm representation.

3.1.1. Periodogram Calculations

After the mel-frequency sub-band decomposition, the derivative of the energy and the periodogram are calculated. The periodogram is the estimation of the power spectrum of a random signal, that is, it looks for the periodicities of any input signal. The periodogram of a discrete signal $x[n]$ is defined as:

$$I(\omega) = \frac{1}{LU} \sum_{m=-(L-1)}^{L-1} c_{vv}[m] e^{-j\omega n} \qquad (1)$$

where L is the length of the segment, U is a normalization factor, $c_{vv}[m]$ is defined as $c_{vv}[m] = \sum_{n=0}^{L-1} x[n]w[n]x[n+m]w[n+m]$, $x[n]$ is the input signal and $w[n]$ is the rectangular windowing sequence. Note that, in fact, the periodogram is the Fourier Transform of the aperiodic correlation of a windowed sequence. From a practical point of view, the periodogram is only calculated at discrete frequencies, that is:

$$I(\omega_k) = \frac{1}{LU} \left| V[k] \right|^2 \qquad (2)$$

where $V[k]$ is the DFT of $v[n] = w[n]x[n]$.

The process starts by sampling the input data at $f_s[Hz]$ and decomposing it into different mel-frequency sub-bands. The frequency analysis is performed each $T[s]$ with a length frame $M[s]$. The derivative of the energy is calculated for each frame and sub-band, then we have the RMS' values with a sampling frequency of $f'_s = \frac{1}{T}[Hz]$. All these RMS' values are saved in a buffer with length $L[s]$. Since the periodogram can be calculated as shown in Eq. 2, the FFT of the squared L length sequence is calculated. In case zero padding is needed, the final sequence has length N, and a unique frequency value can be assigned to each bin. The goal of this method that a unique frecuency value is related to a unique BPM value.

The main advantage of this method is that, for each subband, we have much more information than the information available at the output of a set of resonators tuned at the different BPM typical values. The BPM resolution is higher than other methods, and furthermore, we have not only the BPMs, but all the existing periodicities as well according to different human aspects of music.

3.1.2. Weighted Sum

From now on, we have a rhythmic description of each subband. The *Weighted Sum* of all these new sequences give us a new sequence with the all-band rhythm information. This sum is experimentally weighted, as a function of the style of the input signal. For instance, Pop, Tecno, Rock and, in general, modern commercial music, the information about the rhythm structure is located at low frequencies, that is bass and drums. In Latin Music, rhythmic information is in both low and high frequency sub-bands, that is a wide range timbres from a wide range of percusive instruments. In Classic music, rhythmic information is in any frequency band.

All these experimental values are stored in a vector $s[0, B-1]$, where B is the total number of subbands: If each subband has a periodogram $p_b(0, N-1)$, then the global periodogram can be calculated as:

$$p(0, N-1) = \sum_{b=0}^{B-1} s[b] \cdot p_b(0, N-1) \qquad (3)$$

If the system is used in a context where the genre is a given information, we can emphasize some sub-band rhythm information over the others.

3.1.3. The Rhythm Domain

All this process (periodogram calculations and weighted sum) is what we call *Rhythm Transformation*. For clarity, after the *Rhythm Transformation* calculations, we will work in a so called *Rhythm Domain*[1]. The rhythm domain is a periodic BPM vs. Amplitude coordinate system. The whole process for the rhythm transformation is shown in Fig. 1.

3.2. Rhythm Domain visualization

Which information is available from data in the *Rhythm Domain*? The BPM information can be found as the greatest

[1] We use *Rhythm Transformation* and *Rhythm Domain* nomenclature for clarity, but we are not describing a real transformation from a mathematical point of view.

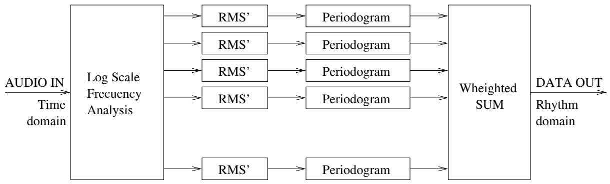


Fig. 1. Block diagram for *Rhythm Transformation*

common divisor for all the representative peaks. That is true because the beat can be defined as the common periodicity of the energy peaks for all the instruments in a song. For BPM detection, any peak detection algorithm across data in rhythm domain data can be used.

But the major advantage of this representation is that it gives us some time domain information too. In rhythm domain, weak beats appear at higher BPM than strong beats. Furthermore, in time domain, weak beats appear later than strong beats too. This correspondence allows us to interpret data in rhythm domain *as* data in time domain. This is what we call *duality* from data in rhythm domain and time domain.

Assuming this duality, we can easily deduce the time signature from audio data. Data between two higher peaks can be interpreted as the distribution of the beats in a bar. If data between two maximum peaks is divided by twos, we assume we are in a simple metre, as we can see in Fig.2(a) and Fig. 2(b). If data between two maximum peaks is divided by threes, we assume we are in a compound metre, as it is shown in Fig. 2(c) and Fig. 2(d). In the other hand, if data in a simple or compound bar is sub-divided by twos, we assume we are in duple metre, as it is shown in Fig. 2(a) and 2(c), and if this data is sub-divided by threes, we assume we are in a triple metre, as it is shown in Fig. 2(b) and 2(d). Finally, in Fig. 2(e) we can see how a simple duple metre sub-divided by twos tells us that this song has swing (assuming swing structure as a dotted quarter-note and a eight-note).

3.3. Beatedness

The *beatedness* calculation is just another application by using data in *Rhythm domain* This concept was introduced by Foote et al in [9] and evaluated by Tzanetakis et al in [10]. The Beatedness is a measure of how strong are the beats in a music piece. In our case, we compute the beatedness as the Spectral Flatness of the sequence in the rhythm domain. Spectral Flatness is a measure of the tonality components in a given spectrum, and it is defined as:

$$SF_{dB} = 10 \cdot \log \frac{G_m}{A_m} \qquad (4)$$

where G_m and A_m are the geometric an arithmetic mean values from all the bins of the Fourier Transform of the signal, respectively.

High beatedness values are due to very rhythmic compositions, this is Dance, Pop, some classical pieces and so on. Low beatedness values are due to non rhytmical compositions, this is some Jazz Solo, classical music, speech...

4. EXPERIMENT

4.1. Description

In our case, the input signal is sampled at $f_s = 22050[Hz]$, the frames are set up to $100[ms]$ long and the hop-size is $10[ms]$. After the frequency analysis, the derivative of energy is calculated and the results for each sub-band are saved in a buffer with a length of $6000[ms]$ and hop-size of $500[ms]$ (The system must be able to recognize a whole bar for a $4/4$ metre at $40bpm$. Then, the frame length must be up to $6[s]$). The BPM value, Rhythm Transformation and Beatedness are computed each $0.5[s]$.

4.2. Results

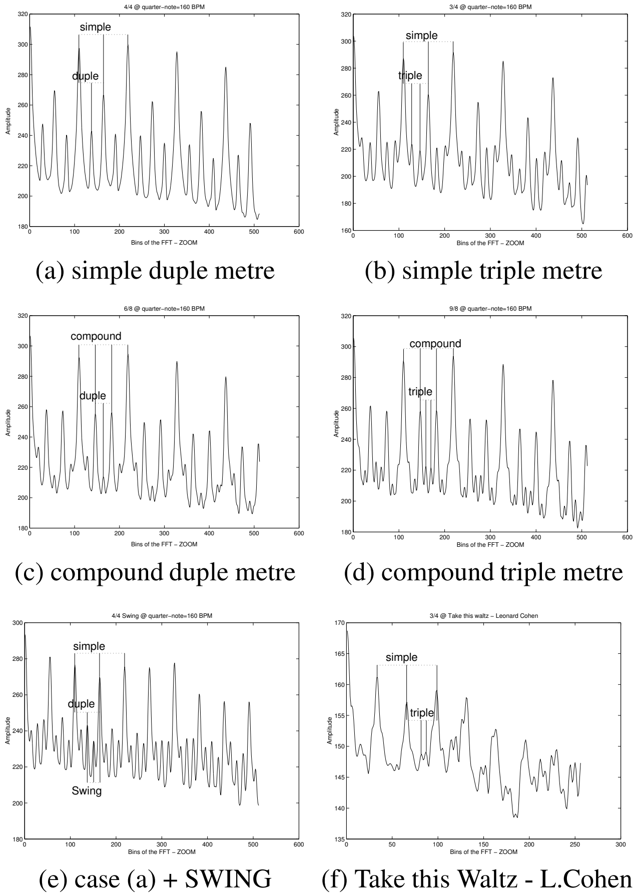
Some BPM & Beatedness measures for different musical genres are shown in Table 4.2. All these measures belong to one frame of "No Gravity" by DJ Session One for Dance music, "Whenever,Wherever" by Shakira for Pop music, "Falling" by Alicia Keys for Soul music, "Summertime" by Gershwin for Jazz music, "Canon" by Pachelbel for Classic music and one minute of radio recording for voice. Note

Genre	BPM_{mean}	BPM_{var}	Beatedness
Dance	140	0.49	3.92
Pop	108	0.71	5.23
Soul	96	0.43	3.48
Jazz	132	1.26	3.54
Classic	90	32	0.95
Voice	66	46.9	0.36

Table 1. BPM and Beatedness for different musical genres

that in *Dance*, *Pop* and *Soul* music, the the system shows us a quite low BPM_{var}. That means that the BPM measure is successful. Not the same for *Classic* and *Voice*. But it's not an error: Do some excerpts of classic music or speech have any tempo? About the Beatedness, high levels are due to rhythmic music, not for *Classic* music or *Voice*.

In Fig. 2, we can see the rhythm domain information for some audio data. Five examples are shown, according to the analysis of one frame of different time signatures: 4/4, 3/4, 6/8 and 9/8 and a swing 4/4 metre. All these examples (available in `http://www.iua.upf.es/~eguaus`) are based on polyphonic MIDI generated audio data. Finally, a real case for 3/4 time signature ("Take this Waltz" by



(a) simple duple metre (b) simple triple metre

(c) compound duple metre (d) compound triple metre

(e) case (a) + SWING (f) Take this Waltz - L.Cohen

Fig. 2. Examples of data in Rhythm Domain for different cases

Leonard Cohen) is shown too. Note that the beats are not equally distributed between the two major peaks because the song does not equally distribute the beats through the bar.

4.3. Limitations

Sometimes, tempo is only defined by pitch variations. As we have seen in Sec. 4.2, the system doesn't work well for those cases with non-attack instruments. Futhermore, the downbeat confusion appears. In this case, a predictive interframe analysis of data in rhythm domain could help us. Finally, we are limited by FFT resolution. For low BPM values, the periodicity is low, then the subdivisions by twos or threes will be much closer than the distance between two bins. The periodogram calculation by using Wavelets instead of the Fourier Transform could be a possible solution.

5. CONCLUSIONS

A new rhythmic representation system has been presented in this paper. In fact, the *Rhythm Transformation* is just a tool that could help us in many different applications. Time signature detection and beatedness calculations are only two of them. As data in *Rhythm Domain* has much more information than a BPM histogram, a lot of musical information can be extracted, as we have shown in Sec.4.2. Some other applications could be found in a Speech/Music Discrimination systems, or in automatic genre classification. At this point, let the imagination fly away.

6. REFERENCES

[1] Stephen Handel, *Listening: An Introduction to the Perception of Auditory Events*, The MIT Press, 2 edition, 1991.

[2] Eric D. Scheirer, "Tempo and Beat Analysis of Acoustical Musical Signals," *J. Acoust. Soc. Am.*, vol. 103, no. 1, pp. 558–601, Jan 1998.

[3] Masataka Goto and Yoichi Muraoka, "A Beat Tracking System for Acoustic Signals of Music," in *Proc. ACM Multimedia*. ACM, 1994.

[4] J. Foote and S. Uchihashi, "The Beat Spectrum: A New Approach to Rhythm Analisis," in *Proc. International Conference on Multimedia and Expo*, 2001.

[5] G. Tzanetakis; G. Essl and P. Cook, "Automatic Musical Genre Classification of Audio Signals," in *Proceedings ISMIR*, 2001.

[6] J. Laroche, "Estimating Tempo, Swing and Beat Locations in Audio Signals," in *Proc. WASPAA*, 2001, pp. 135–139.

[7] Lambrou T.; Kudumakis P.; Speller R.; Sandler M.; Linney A., "Classification of audio signals using statistical features on time and wavelet transform domains," in *Proc. ICASSP, IEEE*, 1998, vol. 6, pp. 3621–3624.

[8] G. Tzanetakis and P. Cook, "Musical Genre Classification of Audio Signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, July 2002.

[9] J. Foote; M. Cooper and U. Nam, "Audio Retrieval by Rhythmic Similarity," in *Proc. ISMIR*, 2002.

[10] G. Tzanetakis; G. Essl and P. Cook, "Human perception and computer extraction of musical beat strength," in *Proc. DAFx-02*, September 2002.