EVALUATION OF MULTIRESOLUTION REPRESENTATIONS OF MUSICAL RHYTHM[*]

Leigh M. Smith and Henkjan Honing
Music Cognition Group
ILLC / Universiteit van Amsterdam
lsmith@science.uva.nl, www.musiccognition.nl

ABSTRACT

A dynamic representation of musical rhythm, the multiresolution analysis using the continuous wavelet transform (CWT), is evaluated using a dataset of the interonset intervals of 105 national anthem rhythms. This representation decomposes the temporal structure of a musical rhythm into time varying frequency components in the rhythmic frequency range (sample rate of 200Hz). Evidence is presented that the beat (typically quarter-note or crochet) and the bar (measure) durations of each rhythm are revealed by this transform. Such evidence suggests that the pattern of time intervals, when analyzed with the CWT, function as features that are used in the process of forming a metrical interpretation. Since the CWT is an invertible transform of the interonset intervals in each rhythm, this result is interpreted as setting a minimum capability of discrimination that *any* perceptual model of beat or meter can achieve. It indicates that a bottom-up, data-oriented process (or a non-cognitive model) is able to reveal durations which match metrical structure from realistic musical examples. This then characterises the data and behaviour of a top-down cognitive model which must interact with the bottom-up process.

1 INTRODUCTION

How does a cognitive structure such as musical meter emerge from exposure to — and consequently, perception of — temporal structure? An ongoing debate continues about the degree to which rhythmic grouping is determined from the signal (data-oriented approach) [13, 6], versus the contribution that cognitive processing (mental model approach) [1] makes to interpretation.

This issue becomes of prime concern when considering how grouping structures develop in computational models. Can they be learned from mere exposure, or is an explicit cognitive process required to capture these structures from a series of musical examples? From another perspective, would a suitably effective general machine learning algorithm be *sufficient* to extract musical structure from the temporal structure of the musical training set? A more nuanced question may posit where the divide between a top-down expectation process and a bottom-up perceptual process lies.

This paper determines that dividing point by taking a non-cognitive approach in evaluating an existing data set of musical rhythms using a representation and visualisation device, the continuous wavelet transform. The data set consists of inter-onset intervals (IOIs) taken from score representations, together with the annotated bar (measure) and beat (quarter note) periods. These periods were tested whether they were present in a time-frequency representation of the rhythms.

The transform which produces a time-frequency representation of rhythm is described in the next section. The method of evaluation and outcomes are described thereafter.

2 THE WAVELET TRANSFORM

Multiresolution representations of rhythm have been demonstrated to reveal periodicities in the temporal structure of onsets [14, 8, 11, 9]. The continuous wavelet transform (CWT) [5, 7] decomposes a time t varying signal $s(t)$ onto scaled and translated versions of a *mother-wavelet* $g(t)$,

$$W_{b,a} = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(\tau) \cdot \bar{g}(\frac{\tau - b}{a})\, d\tau,\ a > 0, \qquad (1)$$

where $\bar{g}(t)$ is the complex conjugate and a is the scale parameter, controlling the dilation of the window function, effectively stretching the window geometrically over time. The translation parameter b centers the window in the time domain. The geometric scale gives the wavelet transform a "zooming" capability over a logarithmic frequency range, such that high frequencies are localised by the window over short time scales, and low frequencies are localised over longer time scales. The CWT indicated in Equation 1 is a scaled and translated instance from a bank of an infinite number of constant relative bandwidth (Q) filters. For a discrete implementation, a sufficient density of scales (a) or "voices" per octave is required.

Morlet and Grossmann's mother-wavelet [4] for $g(t)$ is a scaled complex Gabor function,

$$g(t) = e^{-t^2/2} \cdot e^{i2\pi\omega_0 t}, \qquad (2)$$

where ω_0 is the frequency of the mother-wavelet before it is scaled. Choices of $\omega_0 \geq \pi\sqrt{2/\ln 2}$ will be close to a zero

[*] This research was realized in the context of the EmCAP (Emergent Cognition through Active Perception) project funded by the European Commission (FP6-IST, contract 013123).

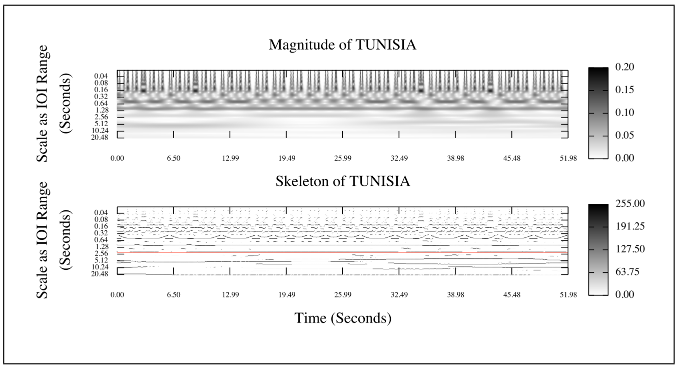

Figure 1. Plots of the magnitude and skeleton (with bar ridge highlighted) of the scaleogram of the Tunisian national anthem rhythm.

mean [5], $\omega_0 = 6.2$ in this application. The Gaussian envelope over the complex exponential provides the best possible simultaneous time/frequency localisation [4], respecting the Heisenberg uncertainty relation. This ensures that all short term periodicities contained in the rhythm will be captured in the analysis. The time domain of $s(t)$ which can influence the wavelet output W_{b_0,a_0} at the point b_0, a_0 is an inverted logarithmic cone with its vertex at b_0, a_0, equally extending bidirectionally in time. Where impulses fall within the time extent of a point, W_{b_0,a_0} will return a high energy value.

By the "progressive" nature of Equation 2 [4, 5], the real and imaginary components of $W_{b,a}$ are the Hilbert transform of each other. These can be computed as magnitude and phase components and then plotted on a "scaleogram" and "phasogram" (upper plot of Figure 1) respectively. The invertibility of the CWT, like the Fourier transform, allows considering the time-frequency domain as a visualisation of the data analysed.

Combinations of the magnitude and phase components can be reduced to time-frequency *ridges* which minimally describe the time varying frequency components in the signal, known collectively as a *skeleton* [12, 9, 10] (lower plot of Figure 1). When applied to musical rhythm, a ridge is an oscillation at a rhythmic frequency, over a period of time, incorporating rubato. Ridges function as beat periods of a rhythm that are prominent and, for example, can serve as the rate that listeners tap or otherwise attend to a musical rhythm. For each rhythm, its skeleton then represents the entire set of beat periods available to a listener to attend to.

3 METHOD

3.1 Anthem Rhythms

While a key benefit of the multiresolution representation of rhythm is a unified representation of both expressive timing and score timing, evaluation in this paper only concerned rhythms represented in score times, lacking expression, each interval being an integer multiple of the minimum duration, typically a 16th note (semi-quaver).

All rhythms were taken from the National Anthem Collection data-set, consisting of 105 rhythms, annotated with score bar (measure) and beat (quarter note) periods [1]. Since these rhythms were taken from scores, the rhythms were all scaled to a single constant tempo of 120 BPM without expressive timing. With a sampling rate of 200Hz, a quarter note was therefore 100 samples (0.5 seconds). This tempo was chosen to minimize the influence of tempo scaling on perception of beat, since it falls close to spontaneous and maximally sensitive rhythm rates [3].

The anthem rhythms ranged in length, and were limited to a maximum of 16384 samples long to restrict the rhythm to a dyadic length to minimize padding the signal and bound the computation time. This translated to 81.92 seconds of performance at 120 BPM with a 200Hz sampling rate, constituting 163 quarter note beats, sufficient to establish a regular grouping and also allow musically typical rhythmic variation. The majority of rhythms were between 40.96 and 81.92 seconds (8192 and 16384 samples) long.

Matching Zaanen's method [15], reduced length rhythms were also analysed, limited to 48 16th (semi-quaver) intervals, comprising the first six seconds of the rhythm if played at 120 BPM, 3 bars at $\frac{4}{4}$. This was done to determine if the extended length of the rhythms were overly biasing the evaluation towards metrical periods, or conversely, if variation over the rhythm was dissipating the spectral peaks.

3.2 Ridge Presence

The average ridge presence vector $\hat{P}$ is the relative frequency of occurrence of a ridge at each dilation scale a, averaged across all rhythms J of a given meter

$$\hat{P} = \sum_{j=0}^{J-1} \frac{P}{J}. \tag{3}$$

The ridge presence vector P of a single rhythm is calculated by summing ridge scales a across the rhythm and normalising for the duration B of each rhythm

$$P_a = \sum_{b=0}^{B-1} \frac{r(W_{b,a})}{B}, \tag{4}$$

where $r()$ is the normalised ridge peak function, derived from the magnitude maxima, local phase congruency and stationary phase measures of each wavelet coefficient $W_{b,a}$, described in detail in [9, 10].

Since the mother-wavelet has a Gaussian profile, and due to the practical limit of the number of voices per octave able to be computed, a single dilation scale will not always perfectly match the period of a signal. For example, with 16 voices per octave, a scale of 106 will correspond to a signal period of 397 samples, the closest scale to a bar period of 400 samples). Therefore the energy will be spread across more than one adjacent scale. To account for this, when measuring the presence of a particular period (for example, bar and beat), the two scales immediately adjacent to the scale a

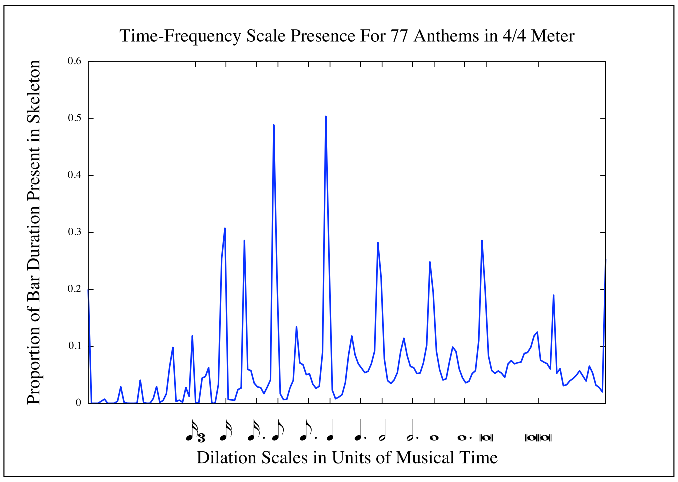

Figure 2. Mean frequency of occurrence of ridges in the skeletons of anthem rhythms in $\frac{4}{4}$ meter. The scales corresponding to common note durations including the bar (semibreve) period are prominent.

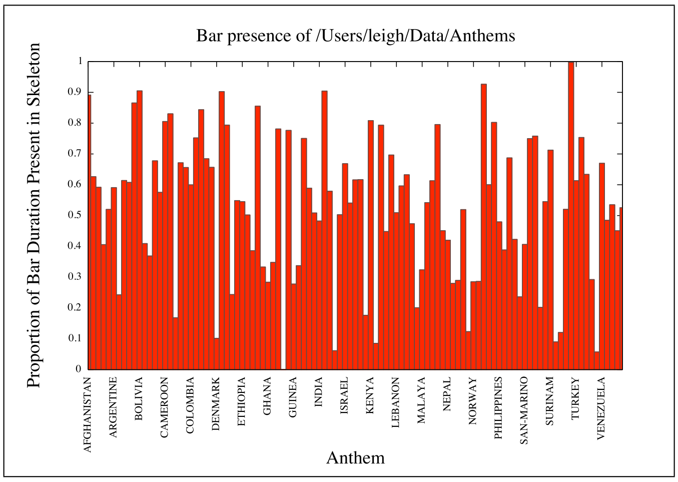

Figure 3. Proportion of bar (measure) periods present in the skeletons of 105 anthem rhythms (long excerpts).

closest to the period examined were also assessed as representing the period. This was computed by "or"ing the three scales together:

$$R_{b,a} = \begin{cases} 1 & \text{if } r(W_{b,a}) + r(W_{b,a-1}) + r(W_{b,a+1}) > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

In the examples tested, this represents an error of ±16 samples for the bar duration in $\frac{4}{4}$ meter rhythms, less than the minimum duration in the rhythms (a 16th note = 25 samples duration). The measure of relative "presence" of the time-frequency scale in the scaleogram was computed as:

$$p = \sum_{b=0}^{B-1} \frac{R_{b,a}}{B}, \tag{6}$$

where B is again the duration of the rhythm and it's respective scaleogram. Thus p ranges from 0 (no ridge around a anywhere in the scaleogram) to 1 (there is energy at every sample along the scale a).

4 EVALUATION

4.1 Bar Presence

Figure 2 illustrates how well the time-frequency representation of different anthems expose the period of the notated bar. The figure displays average relative frequency of occurrence of ridges $\hat{P}$ (Equation 3) for $\frac{4}{4}$ meter, $J = 77$. Each dilation scale a is displayed in musical units and the periods of notes (16th, 8th, quarter notes) are apparent, as is the period of the $\frac{4}{4}$ bar, a whole note (semibreve). The bar period appears due to the decomposition of the temporal structure of the IOIs into rhythmic frequency components.

The stability of the bar duration can be seen by considering how commonly it appears in the skeleton of each rhythm.

	Short (6.0 sec)	Long (81.92 sec)
Bar	0.667 (0.328)	0.528 (0.231)
Beat	0.795 (0.190)	0.763 (0.139)

Table 1. Average and standard deviation of the presence of bar and beat periods in skeletons of the 105 anthem rhythms, for short and long excerpts.

Figure 3 plots the relative presence of the bar duration, calculated by Equation 6. The average and standard deviation of this relative presence of the bar and the beat periods in the skeletons of the anthems is summarised in Table 1. In contrast to the common occurrence of the bar duration in nearly all of the skeletons, only 33 of the 105 anthem rhythms have any occurrences of the bar duration IOI, demonstrating how the decomposition of the CWT reveals the period of the bar.

While a multiresolution representation has been shown to reveal frequency components matching the period of the bar, which is typically the rate that listeners will group the rhythm, the CWT alone is insufficient for a complete theory of perception of rhythm. For example, there is currently no use of rhythmic phase to identify an anacrusis (upbeat). Since the anthem dataset is annotated with upbeats, a future task is to develop and evaluate such an extension of the CWT model.

4.2 Case Studies

Given the lack of expressive timing and dynamic, harmonic and melodic accents, it is surprising that such metrical periods are exposed as well as they are from the temporal structure alone. In most melodies, the melodic structure would contribute to the metrical grouping structure, so some rhythms which do not display a strong bar ridge are to be expected.

An example is the Greenland anthem which has no spectral energy at the bar period ($\frac{4}{4}$, 400 samples, 2.0 seconds). However it does have prominent ridges at periods of the 16th, 8th, quarter, half notes, at 3 quarter notes and 5 quarter notes.

These last two periods arise since the anthem is a dotted eighth and dotted quarter note rhythm and lacking disambiguation by accenting, the common repetition of these durations decomposes onto frequency components of 12 and 20 16ths.

At the other extreme, the Tunisian anthem rhythm has a well defined continuous ridge at the bar duration of 400 samples (2 seconds) as shown in Figure 1. This rhythm is also a dotted rhythm, however it more regularly repeats a dotted rhythmic figure that has a period that matches the bar duration. The Greenland rhythm has several longer intervals, which leads the decomposition of that rhythm to favour the periods of 3 and 5 quarter notes, whereas the Tunisian rhythm decomposes more parsimoniously into 4 quarter notes. The variation in timing across the duration of each anthem produces the range of bar presence measures as reflected in the standard deviation measures in Table 1.

5 CONCLUSION

This paper has evaluated the continuous wavelet transform as a multiresolution representation of musical rhythm [8, 9, 11, 10] against a known dataset. With the decomposition of the rhythm into time-varying frequency components, periods which match intervals of the beat and bar are revealed.

The presence of metrical durations, revealed to be frequency components of the temporal structure of a rhythm, suggests that they might function as cues for meter in an emerging cognitive construct created by active perception. This suggests a sufficiently discriminating bottom-up perceptive mechanism can, potentially, provide features (i.e. ridges) that are then statistically clustered into cognitive features (the current metrical interpretation). That metrical features can appear from signal representations, such as the CWT decomposition, demonstrates this information can be derived from the data itself without an explicit model of rhythmic cognition.

Of course, a bottom-up approach alone does not account for all processes that are required for the perception of meter and rhythm. An additional cognitive model or top-down process is essential in, for example, the disambiguation of the metrical alternatives present in the signal, the identification of ridges which match against cognitive features such as the tactus (the metrical level that is most salient), and perpetuation of a metrical interpretation in the face of syncopation or other contradictory evidence from performed events.

This then suggests that a top down process would function by evaluating the candidate ridges (evidence of time varying rhythmic periodicities) and selectively use a subset at any given time to form an attending strategy. The simplest example strategy is choosing a single ridge that can be clapped to [9, 10]. This selection process need not be exclusionary, enabling several competing candidates to be continuously assessed according to their fit within existing mental schemas and perceptual and performance bounds. It is likely that the bottom-up process is not only ridge formation, but is mediated by a categorisation process which then interacts sequentially with a top-down meter induction process [2, Fig. 3].

A future research task is to compare the performance of human listeners in selecting tactus periods to the decomposition behaviour that the CWT produces and to the bar as notated. Testing on larger, richer datasets that includes tempi, melodic information and compound metrical rhythms is also a future research task.

6 REFERENCES

[1] P. Desain and H. Honing. Computational models of beat induction: The rule-based approach. *Journal of New Music Research*, 28(1):29–42, 1999.

[2] P. Desain and H. Honing. Modeling the effect of meter in rhythmic categorization: Preliminary results. *Japanese Journal of Music Perception and Cognition*, 7(2):145–56, 2001.

[3] P. Fraisse. Rhythm and tempo. In D. Deutsch, editor, *The Psychology of Music*, pages 149–80. Academic Press, New York, 1st edition, 1982.

[4] A. Grossmann, R. Kronland-Martinet, and J. Morlet. Reading and understanding continuous wavelet transforms. In J. Combes, A. Grossmann, and P. Tchamitchian, editors, *Wavelets*, pages 2–20. Springer-Verlag, Berlin, 1989.

[5] M. Holschneider. *Wavelets: An Analysis Tool*. Clarendon Press, 1995. 423 p.

[6] A. P. Klapuri, A. J. Eronen, and J. T. Astola. Analysis of the meter of acoustic musical signals. *IEEE Transactions on Audio, Speech and Language Processing*, 14(1):342–55, 2006.

[7] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998. 577p.

[8] L. M. Smith. Modelling rhythm perception by continuous time-frequency analysis. In *Proceedings of the International Computer Music Conference*, pages 392–5. International Computer Music Association, 1996.

[9] L. M. Smith. *A Multiresolution Time-Frequency Analysis and Interpretation of Musical Rhythm*. PhD thesis, Department of Computer Science, University of Western Australia, July 1999.

[10] L. M. Smith and H. Honing. Time-frequency representation of musical rhythm by continuous wavelets. *(submitted)*, 2007.

[11] L. M. Smith and P. Kovesi. A continuous time-frequency approach to representing rhythmic strata. In *Proceedings of the Fourth International Conference on Music Perception and Cognition*, pages 197–202, Montreal, Quebec, August 1996. Faculty of Music, McGill University.

[12] P. Tchamitchian and B. Torrésani. Ridge and skeleton extraction from the wavelet transform. In M. B. Ruskai, editor, *Wavelets and Their Applications*, pages 123–51. Jones and Bartlett Publishers, Boston, Mass., 1992.

[13] N. M. Todd, C. Lee, and D. O'Boyle. A sensorimotor theory of temporal tracking and beat induction. *Psychological Research*, 66(1):26—39, 2002.

[14] N. P. Todd. The auditory "primal sketch": A multiscale model of rhythmic grouping. *Journal of New Music Research*, 23(1):25–70, 1994.

[15] M. van Zaanen, R. Bod, and H. Honing. A memory-based approach to meter induction. In *Proceedings of the ESCOM*, pages 250–253, 2003.