Towards Rhythmic Analysis of Human Motion using Acceleration-Onset Times

Eric Lee, Urs Enke and Jan Borchers
Media Computing Group
RWTH Aachen University
52056 Aachen, Germany
{eric, borchers}@cs.rwth-aachen.de
urs.enke@rwth-aachen.de

Leo de Jong
MultiPRO
Kromhoutlaan 1
2033 WJ Haarlem, The Netherlands
leo@multipro.demon.nl

ABSTRACT

We present a system for rhythmic analysis of human motion in real-time. Using a combination of both spectral (Fourier) and spatial analysis of onsets, we are able to extract repeating rhythmic patterns from data collected using accelerometers. These extracted rhythmic patterns show the relative magnitudes of accentuated movements and their spacing in time. Inspired by previous work in automatic beat detection of audio recordings, we designed our algorithms to be robust to changes in timing using multiple analysis techniques and methods for sensor fusion, filtering and clustering. We tested our system using a limited set of movements, as well as dance movements collected from a professional, both with promising results.

Keywords

rhythm analysis, dance movement analysis, onset analysis

1. INTRODUCTION

Few would dispute the essential connection between rhythm and music – some researchers have even claimed that music is the "rhythmization of sound" [11]. Regardless, rhythm indeed plays a strong role in our perception and interpretation of music. It is also one of the key components that form the symbiotic relationship between dance and music that dates back to prehistoric times; body movements and music are closely linked in a dynamic relationship between acting and listening, cause and effect.

Unfortunately, there has been little work studying this connection between rhythm, dance and music in designing new musical interfaces. Existing systems for creating music from gestures often employ spatial mappings [9, 13] with little consideration for the temporal aspect of tempo or rhythm. The work presented in this paper is part of a larger project aimed at studying rhythm in the context of music and dance. One of the primary research challenges is to extract and identify rhythmic information from dance movements. Potential applications of this work include a system that allows dancers to directly influence the rhythm of pre-recorded music, automatic classification of a dance performance based on the extracted rhythm, and rhythm-based approaches to sonification of dance.

Extracting rhythmic information from arbitrary movements is a complex endeavor, and in this work, we present an approach to tackle a subset of this task – extracting rhythmic patterns from movement of a single limb using accelerometers in real-time. We will begin by clarifying the terminology that will be used throughout the paper, followed by a discussion of related work that inspired our algorithm design and implementation. We conclude with the results of some informal evaluation, both with test data and data collected from professional dancers.

NIME07, June 7–9, 2007, New York, NY

2. TERMINOLOGY

Despite the ubiquity of the term "rhythm", its exact definition remains a matter of some controversy [16]. Guedes, as part of his work on studying dance and music [10], studied various views of the term. He notes that the rhythmic perception when watching a dancer is strongly determined by accompanying music and difficult to attain in silence. Moreover, a dance often cannot capture all of music's rhythmical elements for physical reasons. However, the spatial elements present in dance (multiple limbs moving, stepping patterns) may provide rhythmic cues. In his work, he adopted Parncutt's definition that "a musical rhythm is an acoustic sequence evoking a sensation of pulse" [14].

Another definition, by Dowling and Harwood, is "a temporally extended pattern of durational and accentual relationships" [4]. This definition seems to be appropriate when talking about rhythm and music, and in this paper we adopt a similar definition, where we will refer to a "rhythm pattern" as "a repeating series of accentuations of impulses separated by time intervals". In a music setting, impulses would be notes and their accentuations could be determined by their volume, and in a dance setting, they could be movements with their respective maximal momentum. A graphical representation of an example rhythm pattern is shown in Figure 1. We will also use the term "beat" to refer to a single element within the pattern. A rhythm pattern as we have defined it, then, consists of multiple beats of varying magnitudes spaced roughly evenly apart.

3. RELATED WORK

Our work is inspired by two main areas of research: analyzing dance movements, and analyzing musical rhythm. Due to space constraints, we are only able to provide a short overview of these works here; a more exhaustive review can be found in [6].

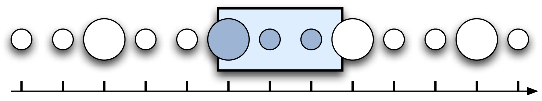

Figure 1: An example representation of a rhythm. Each circle along the horizontal time axis represents an impulse, its size signifying the impulse's magnitude. As the pattern *strong–weak–weak* is repeating, it can be called a rhythm.

3.1 Dance Movement Analysis

Paradiso et al.'s *DanceShoe* [12], in its most current incarnation, produces readings of 16 different parameters, including those of differently tilted accelerometers and pressure sensors at various positions. The goal of their work was to give "improvisational dancers a 'palette' of action-to-sound rules and relationships" of varying nature, using acceleration sensors either as a potentiometer (e.g., correlating a tilt angle with an instrument's pitch) or as a binary switch (shock movements causing individual sounds or playing a drumroll while doing a handstand).

Feldmeier [7] also uses accelerometers for creating auditory feedback from multiple dancers, and employs Fourier analysis to efficiently analyze pulses from the dancers, triggered when the acceleration exceeded a certain threshold. Among the information that could be obtained from this analysis was tempo.

Other approaches include using floor sensors [18] and capturing video of dancers [10]; Guedes, in particular, shows how a dance performance's tempo can be determined by examining brightness changes between video frames.

In the works described above, the temporal information derived from dance movements is limited (e.g., only tempo), if at all.

3.2 Musical Rhythm Analysis

The problem of extracting musical rhythm from accelerometer data is, in some ways, similar to analyzing audio recordings. In such analysis, there is often first a conversion to a symbolic representation, from which the desired information is extracted, although there have been attempts to process the raw audio data directly using comb filters [15], auto-correlation [8], or Fourier analysis [17].

Cemgil and Kappen [2] employed various probabilistic methods for quantization and tempo tracking built around Bayes' theorem. Brown employed auto-correlation to determine the musical meter of a score [1]; while her method worked well with perfectly timed files, she was not as successful with performance recordings which, by nature, carry a certain level of imprecision in the timing. Some current research has focused on rule-based beat induction; Eck [5] analyzed several rule-based approaches, and compared the beats they identified on a meter grid with what human listeners intuitively tapped to. The generalized detection algorithm assumed the meter hierarchy to be known, however.

A popular method of inferring meter information from symbolic sequences of impulses is to analyze the intervals between note onsets, commonly referred to as inter-onset intervals (IOIs). A statistical analysis of IOIs could give an indication of that sequence's beat interval and measure length. Dixon [3] used clustering to determine the tempo of a recording, and Seppänen [16] used a histogram approach.

3.3 Intended Contribution

There are several aspects in which this work builds upon the works presented above. The first is the application of rhythmic extraction to motion data. Among those dealing with dance inputs, only Guedes was interested in the extraction of rhythmic information; however, his analysis was based on video data rather than accelerometer data. Accelerometer data has the advantage of both finer spatial granularity and higher temporal resolution.

Current systems that extract tempo and meter information focus on music recordings, and often exploit additional information not present in motion data, such as harmonic and tonal hints, in their analysis. Moreover, many of these algorithms do not work in real-time. In this work, we also produce a representation of the actual rhythm – most of the mentioned works concentrate on finding the tempo or limit other output to low-level information such as, in Guedes' case, the tempo's harmonics. Finally, our algorithms have been designed to be robust to data collected from expressive performances, where the rhythm timing varies as the piece progresses.

4. DESIGN

The design of our algorithm is inspired by the musical rhythm analysis literature outlined above. We incorporate two types of sensor data analysis: interval and frequency. Interval analysis has the benefit of low latency (an impulse can be processed and contribute to an updated result as soon as it is detected); frequency analysis, on the other hand, has the benefit of being more robust in the presence of noise. Figure 2 shows a block diagram illustrating our approach to rhythmic analysis of accelerometer data. In the following sections, we will outline in detail each of the steps: movement detection, interval analysis, frequency analysis, data fusion, impulse folding, and impulse clustering.

4.1 Movement Detection

Movement detection takes each channel of sensor data and extracts an impulse sequence.

The accelerometer data for a single downwards movement is shown in Figure 3. There are two opposing pulses: an acceleration pulse when the movement starts, and a deceleration pulse when the movement is stopped. This pulse pattern can also be clearly seen for regular up-down movements, and we have a sequence of acceleration and counter-acceleration pulses (see Figure 4). The intensity of the movements also corresponds to the size of the pulses.

Unfortunately, the pattern becomes increasingly complex for unrestricted movements (see Figure 5). This "noise" can be attributed to a variety of causes: involuntary twitches, interference of various limbs' and other body parts' movements or the influence of Earth's gravity. Fourier analysis methods help in extracting some of the parameters needed for the rhythmic analysis in the presence of such noise, and are discussed later in this paper.

To analyze movements of interest, we extract three parameters from the sensor data for each impulse (see Figure 6):

- **Timestamp**: Determining a suitable timestamp $\tau(m)$ for a movement m is largely dependent on the type of movement. For some movements, such as hand-clapping, the time of interest is clearly when the two hands come together; for others, such as ballroom dancing, the mapping between the musical rhythm and the swings and turns of the movements is less clear. For the purposes of this work, we arbitrarily decided to use the midpoint between the start and maximum point of acceleration as the timestamp of the movement, and leave the problem of studying the exact temporal mapping of movements for specific dance genres to future work. We believe the results of such studies could be easily incorporated into our algorithms as a function of the impulse spread.

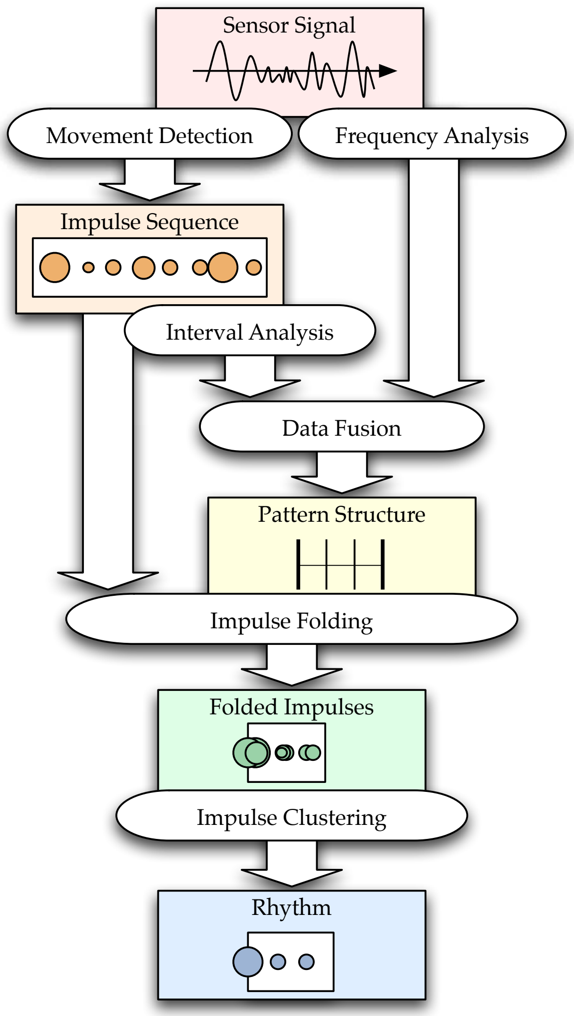

Figure 2: Block diagram of our rhythmic analysis system. The raw signal data from the accelerometers is processed using movement detection followed by interval analysis, and a second path using frequency analysis. The results are combined using data fusion, followed by impulse folding and clustering to obtain the final result.

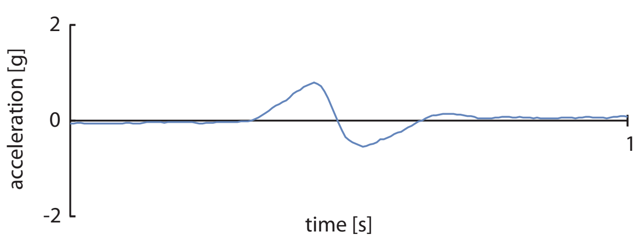

Figure 3: Acceleration graph of a sudden drop of the sensor. The first pulse is a result of the increase in acceleration, and the second pulse a deceleration to stop the movement.

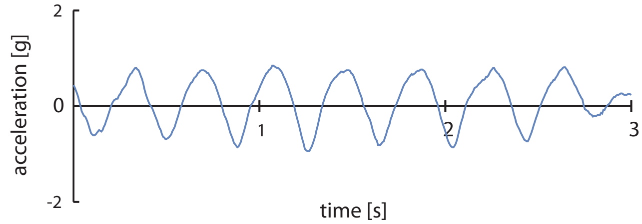

Figure 4: Acceleration graph of continuous up-down movements. There is no explicit deceleration pulse, and pulses are balanced in both directions.

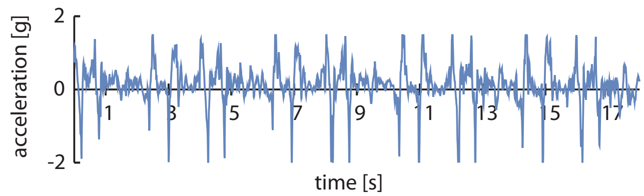

Figure 5: Acceleration graphs of the right shin of a *Cha-cha-chá* dancer. The movements are not equidistant because the dancer alternated step sequences between the left and right feet.

- **Magnitude**: We use the area of the acceleration pulse as a measurement of the impulse magnitude $\mathrm{mag}(m)$, corresponding to the intensity of the movement; the area is computed using a sum of all sensor values over the impulse spread. We experimented with other schemes, such as using an pulse's amplitude relative to the current average activity level; such a scheme would, in theory, extract only the most prominent movements from the data. However, we found that it also unrealistically assigns more importance to tiny movements in almost motionless sensors.

- **Spread**: The impulse spread $\Delta(m)$ is defined by where the acceleration pulse crosses the zero point on either side. Combined with the area, it is thus possible to distinguish between slow, soft movements and fast, sudden movements.

In the subsequent analysis it is desirable to differentiate acceleration pulses (which start a movement) from deceleration pulses (which stop a movement). However, as can be seen in Figures 3 and 4, such a distinction is not possible; to address this, we split the

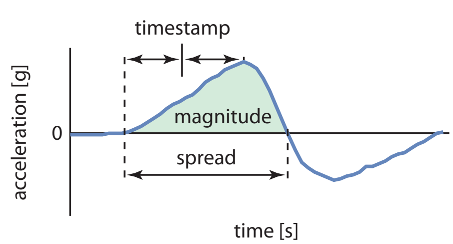

Figure 6: The three parameters extracted for movement detection. The midpoint between the start of the pulse and its maximum point of acceleration is used for the timestamp. The magnitude is simply the area of the acceleration pulse, and the spread is calculated from the two closest zero-crossings of the acceleration graph.

positive and negative pulses into two streams. We found that deceleration pulses usually have lower magnitude, which is accounted for in the data fusion stage.

4.2 Interval Analysis

The aim of the interval analysis is to find an interval between individual events in a repeating rhythmic pattern. We define the distance between two rhythmic events as a *beat interval* and the length of a repeating rhythmic pattern a *pattern length*.

We first compute a set of weighted *inter-impulse intervals* (IIIs). IIIs are analogous to the inter-onset intervals defined for rhythm analysis of music; we use the interval between the impulse timestamps computed in the previous step:

$$III = \tau(m_1) - \tau(m_2) \tag{1}$$

These intervals are assigned a weight, which is the minimum of the two magnitudes:

$$\mathrm{mag}(III) = \min\left(\mathrm{mag}(m_1), \mathrm{mag}(m_2)\right) \tag{2}$$

The inter-impulse interval spread also provides a measurement of uncertainty:

$$\Delta(III) = \frac{\Delta(m_1) + \Delta(m_2)}{2} \tag{3}$$

All possible inter-impulse intervals for the last two seconds of data are then accumulated into a histogram; the histogram has the interval size on the horizontal axis and the magnitude on the vertical axis (see Figure 7). The interval size is quantized in 20 ms intervals, also referred to as "buckets". To account for the spread representing the uncertainty, the IIIs are not accumulated as impulses in a single bucket, but as triangles with height $\mathrm{mag}(III)$ and width $\Delta(III)$ (the spread value).

To account for history beyond the last two seconds, and also to guard against erratic data, the histogram is averaged with the previously calculated histogram. This one pole low pass filter technique across histograms was also employed by Seppänen [16] for similar reasons.

From Figure 7, we can see that the pattern length occurs at the maximum peak in the histogram, and the beat interval at the first "significant" peak. When searching for peaks, we use two criteria: a data point is considered a peak if it is larger than its two neighboring buckets on either side. A peak must also be larger than the average magnitude across the entire histogram.

4.3 Frequency Analysis

The frequencies we are interested in extracting from the sensor data are in the range of a few Hertz or less, which requires a time window of a few seconds. This latency makes the results from a pure frequency analysis, in general, unsuitable for real-time. However, we still perform the analysis, and combine it with the results of our interval analysis to increase the reliability of our results. We transform a ten-second time window of sensor data into the spectral domain; the fundamental frequency, then, is our previously defined beat interval. In our current implementation, we first downsample the data by a factor of six, followed by a 256-point Fast Fourier Transform (FFT). We consider a data point to be a peak in the signal spectrum when the amplitude is larger than its two neighboring frequency bins.

4.4 Data Fusion

We require a data fusion scheme to combine the results from both multiple sensors (one for each axis of movement) and analysis

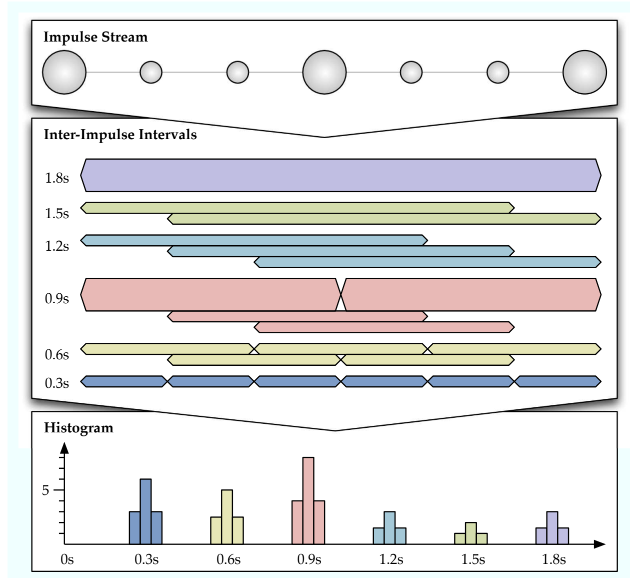

Figure 7: Example of histogram accumulation. The histogram is created using all possible inter-impulse intervals over the last two seconds of data. The colored bars represent the IIIs between the impulses; the thickness of the bars is an indication of their magnitude. To account for uncertainty, a triangle with width equal to the III spread is accumulated into the histogram. The highest peak at 0.9 s is the pattern length; the first peak at 0.3 s is the beat interval.

types (interval and frequency). We use a voting scheme where the results of the analyses are again histogrammed based on the computed beat interval and pattern length. The values with the highest count are then passed to the impulse folding module. We again adopt the one pole low pass filtering technique with previously accumulated histograms here, with the assumption that data sources that were previously reliable remain reliable for the short-term.

4.5 Impulse Folding

With the beat interval and pattern length, we now know, approximately, the length of a repeating rhythm pattern. We use this approximate length to divide the impulse stream into shorter segments and overlay them on top of each other so that they form a repeating pattern.

We assume the first beat of the pattern is the strongest one, and use that to decide where to perform this "folding" operation. The divided segments will be of slightly different length, and to assist the subsequent clustering process, we normalize the length of the impulse segments to the pattern length (see Figure 8).

4.6 Impulse Clustering

In this final step of the algorithm, we look for impulses that are close to each other and combine them into a single impulse; the average of the magnitudes are taken. This produces the repeating pattern shown at the bottom of Figure 2.

5. IMPLEMENTATION

The various modules described in the previous section are implemented as a set of Java classes. We also wrote a set of wrappers to these classes so that they can be used as Max/MSP externals[1]. For the accelerometers, we used prototype hardware designed by

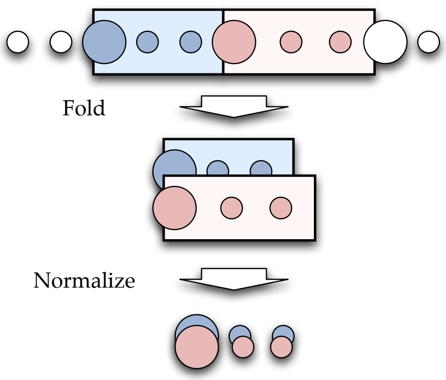

Figure 8: An impulse stream is folded by dividing the impulses into segments based on the calculated pattern length. The individual segments are then normalized to this pattern length to assist in the clustering process.

one of the authors in conjunction with an engineering group in Romania; details of this hardware, which consists of accelerometers communicating with a base unit and supporting software, are unfortunately beyond the scope of this paper. However, our Max/MSP patches communicate with the sensor base unit using Open Sound Control, making it possible to use any type of accelerometer sensor. Figure 9 shows an example of the system running under Max/MSP on Mac OS X.

6. EVALUATION

We tested our algorithms with a variety of rhythm patterns performed by test users. Sensors were attached to the finger or hand, and the rhythm pattern was performed by waving in mid-air with circular gesturing motions to trigger multiple axes of movement (such as the *strong-weak-weak* pattern shown in Figure 9). The system works well for these types of movements; the rhythmic pattern is recognized within 6 seconds, starting from rest. Both the beat interval and the pattern length are correctly reported. Our algorithm can correctly identify patterns with pauses in between, such as the pattern *strong-weak-rest-weak*, although with less reliability than patterns with evenly spaced beats.

We did identify several cases where our algorithm reports partially inaccurate results. In cases where both the first and second beat are performed with roughly equivalent magnitude, such as *strong-strong-weak-weak*, the pattern is folded at the correct point; however, the second beat event is visibly more pronounced than the first. We attribute this to an artifact of the histogramming and clustering inaccuracy.

A second case where the algorithm is problematic in reporting correct results is when the rhythmic pattern contains more than 5 beats. In this case, half of the histograms report that the first impulse is the largest (if only by a small amount), resulting in an incorrect pattern length (see Figure 10). It would appear in this case that the sheer number of pairs one interval apart outweighs the magnitude difference.

We also ran our algorithm through data recorded from a professional *Cha-cha-chá* dancer (see Figure 11). While it is not able to capture the exact *one-two-three-cha-cha* rhythm[2], it was able to correctly identify the pattern length and the accents on the first and

[1]The source code for this project is freely available at `http://media.informatik.rwth-aachen.de/enke.html`.

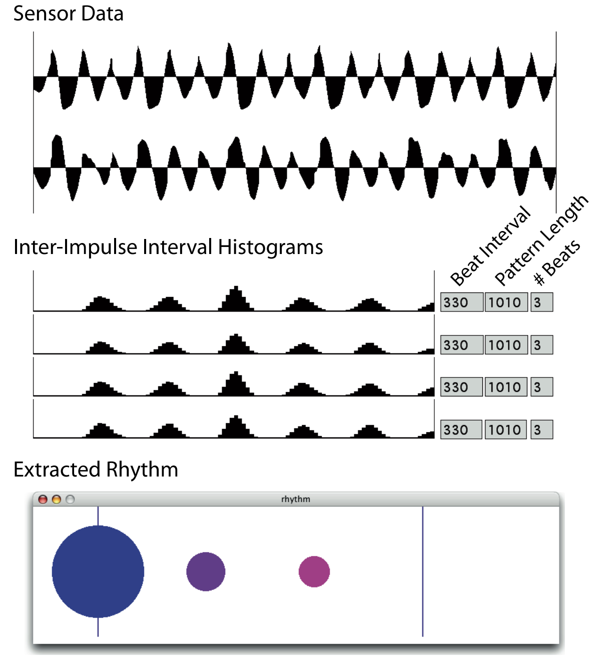

Figure 9: Result of a user gesturing in a 3-beat (*strong-weak-weak*) pattern using circular movements. Such movements trigger accelerometers along two axes, which are in turn split into positive and negative pulses, resulting in four histograms. The three numbers beside each histogram are the beat interval, pattern length, and number of beats per pattern, respectively. The resulting pattern is correctly identified.

third beats.

We hope to address these shortcomings in future work.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we introduced a system for extracting rhythmic patterns from human movements using accelerometer data. Inspired by previous work in automatic beat and tempo detection in musical recordings, our system converts accelerometer data into a stream of impulses from which beat interval and pattern length information are extracted. This information is then used to identify repeating patterns in the rhythmic stream, and output them as strings of

[2]The *Cha-cha-chá* rhythm is also commonly written as *step-step-cha-cha-cha* – however, the last "*cha*" corresponds to the first accented beat.

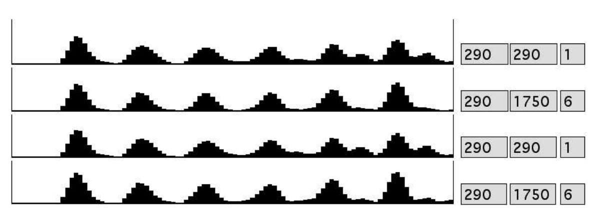

Figure 10: The histograms generated for a 6-beat (*strong-weak-weak-weak-weak-weak*) pattern. Even though the pattern length should be 6, the sixth peak is slightly lower than the first peak in two of the four histograms, and the algorithm incorrectly reports a pattern length of 1.

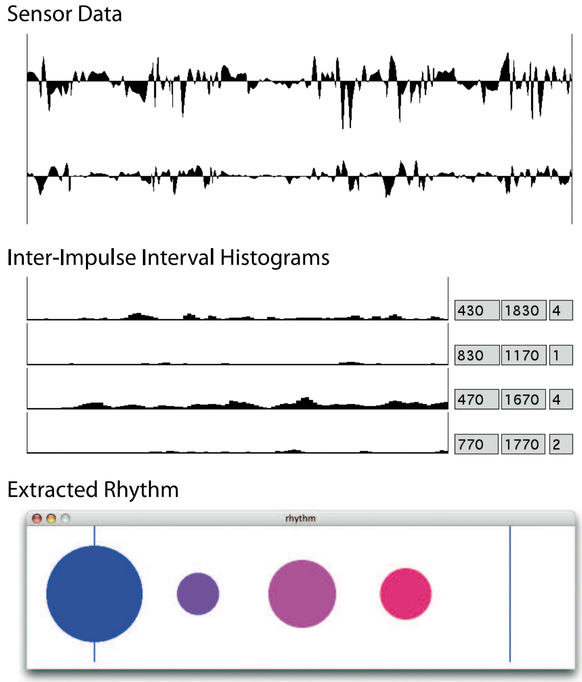

Figure 11: Results of running our algorithm with data collected from a *Cha-cha-chá* dancer. The exact rhythm is not captured, but the correct pattern length is identified, and the algorithm correctly detects the emphasis on the first beat and a smaller emphasis on the third beat.

weighted values and their spacing in time. We use a combination of spatial and spectral algorithms in our analysis. Spatial algorithms have the benefit of lower latency, but are more prone to error; spectral algorithms are more robust, but introduce more latency into the system. A voting mechanism allows us to combine the results of these analyses for increased reliability.

Our current system has been shown to work reliably for well-defined movements from a single limb, and we have identified a number of areas of future work:

- **Multiple sources**: A rhythmic pattern is often created from two limbs with a phase offset, much as a drummer would do, and identifying such patterns would be a natural extension to this work.
- **Alternative analysis methods**: In addition to the spatial and frequency-based analysis methods we would like to explore analyses based on the wavelet transform.
- **Analysis with real dance movements**: Our work focused on showing that such an analysis is viable. As we continue this work, we hope to run more user tests with professional dancers, and correlate their mental models of rhythm with the output from our system.

We hope our work will inspire further studies of the intricate mappings between music, rhythm, and human motion.

8. ACKNOWLEDGEMENTS

We would like to thank Daniel Moga and his colleagues at the Technical University Cluj-Napoca in Romania for their work on the accelerometer hardware. We would also like to thank Martin Neuhäußer, Alena Sudholt, Lorenz Merdian, Niels Gürtler, and Thomas Rath for allowing us to record their dance movements.

9. REFERENCES

[1] J. C. Brown. Determination of the meter of musical scores by autocorrelation. *Journal of the Acoustical Society of America*, 94(4):1953–1957, 1993.

[2] A. T. Cemgil and B. Kappen. Monte carlo methods for tempo tracking and rhythm quantization. *Journal of Artificial Intelligence Research*, 18(1):45–81, 2003.

[3] S. Dixon. Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30(1):39–58, 2001.

[4] W. J. Dowling and D. Harwood. *Music Cognition*. Academic Press, San Diego, USA, 1986.

[5] D. Eck. A positive-evidence model for rhythmical beat induction. *Journal of New Music Research*, 30(2):187–200, 2001.

[6] U. Enke. DanSense: Rhythmic analysis of dance movements using acceleration-onset times. Master's thesis, RWTH Aachen University, Aachen, Germany, September 2006.

[7] M. Feldmeier. Large group musical interaction using disposable wireless motion sensors. Master's thesis, Massachusetts Institute of Technology, Massachusetts, USA, February 2003.

[8] F. Gouyon and P. Herrera. Determination of the meter of musical audio signals: Seeking recurrences in beat segment descriptors. In *114th Convention of the Audio Engineering Society*, Amsterdam, The Netherlands, March 2003.

[9] N. Griffith and M. Fernström. Litefoot - a floor space for recording dance and controlling media. In *Proceedings of the ICMC 1998 International Computer Music Conference*, pages 475–481, Ann Arbor, USA, 1998.

[10] C. Guedes. *Mapping Movement to Musical Rhythm: A Study in Interactive Dance*. PhD thesis, New York University, New York, USA, 2005.

[11] C. Hasty. *Meter As Rhythm*. Oxford University Press, 1997.

[12] J. Paradiso, K.-Y. Hsiao, A. Benbasat, and Z. Teegarden. Design and implementation of expressive footwear. *IBM Systems Journal*, 39(3&4):511–529, 2000.

[13] J. Paradiso, K.-Y. Hsiao, J. Strickon, J. Lifton, and A. Adler. Sensor systems for interactive surfaces. *IBM Systems Journal*, 39(3&4):892–914, 2000.

[14] R. Parncutt. A perceptual model of pulse salience and metrical accent in musical rhythm. *Music Perception*, 11(4):409–464, 1994.

[15] E. D. Scheirer. Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America*, 103(1):588–601, 1998.

[16] J. Seppänen. Computational models of musical meter recognition. Master's thesis, Tampere University of Technology, Tampere, Finland, August 2001.

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

[18] P. Srinivasan, D. Birchfield, G. Qian, and A. Kidanè. A pressure-sensing floor for interactive media applications. In *International Conference on Advances in Computer Entertainment Technology*, pages 278–281, Valencia, Spain, 2005.