Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institut für Elektronik
Prof. G. Tröster
Herbstsemester 2009

Semester Thesis

Who is Dancing?
Recognize Dancers by Means of Acceleration Sensors

Author:
Matthias Flückiger

Readers:
Martin Wirz
Michael Kuhn

Contents

List of Figures

List of Tables

Abstract

This work investigates on identifying dancers and non-dancers by correlating features extracted from audio and acceleration data. Acceleration data was generally caught by one body-worn acceleration sensor per person only.

Two experiments were accomplished to decide about sensor location, to derive common features of acceleration and audio signals, to test promising algorithms and to evaluate the elaborated methods.

In Chapter 1 the set up and the thoughts that led to the configurations of the experiments are explained. Chapter 2 focuses on signal properties of the involved signals and finally presents a method that is able to discriminate dancers and non-dancers accurately.

Introduction

Dancing belongs to one of the most expressive and important forms of non-verbal communication. In [1], Adshead writes: “Dance [. . .] is a sport and art form that generally refers to movement of the body, usually rhythmic and to music”. This definition suggests that body-worn acceleration sensors should be able to detect if somebody is dancing to a given piece of music.

One can imagine an automatic DJ system that could use such an information as a feedback of the overall mood in a club or in a discotheque. Such a system could be able to explore the taste of music of an attending crowd. The automatic DJ would be able to interact with the crowd by generating playlists based on the collective parameter of how many people are dancing to the played back music.

This Semseter Thesis aims at investigating algorithms and methods to characterize dancing people by means of acceleration sensors. The goal is to discriminate dancers and non-dancers by correlating features extracted from acceleration data and the played back music. Only one body-worn acceleration sensor has been used to keep the possibility of applying the presented methods to the use of data captured by mobile phones.

To elaborate and evaluate the algorithms two different experiments had been accomplished and the observed characteristics and similarities of audio and acceleration are presented in this work.

Background

There are two interesting limits concerning the interaction of humans with music. On one side Winter [2] demonstrates that the useful frequency spectrum of human motion lies in the range of 0 to 10 Hz, on the other hand Noorden et al. [3] tell us that frequencies above 4 Hz are irrelevant to human rhythm perception.

Related Work

There are various research projects that deal with dancing and music. Motion is generally caught by a computer vision system or by the use of one or more acceleration sensors. We decided to use one acceleration sensor per person only, to allow to adapt the presented procedures to the use of state-of-the-art mobile phones instead of attached acceleration sensors. For an automatic DJ system the capability to be used with mobile phones is important, because the willingness of visitors to attach sensors before dancing may be rather small.

On one hand there are attempts that investigate the similarity of movements in simultaneous dancing or social interaction [4, 5, 6], on the other hand there are several approaches that concern the interplay of individuals to music [7, 8, 9]. The interaction of a crowd of people with music has found little interest in research so far.

For the study of behavior and context analysis of movements, data is often captured by the use of accelerometers and processed with a correlation measure. Lester et al. [4] present a method which determines whether two devices are carried by the same person or not. For this task they use the coherence function, which is a measure of how much two signals are linearly related at each frequency. The coherence is computed from the raw signal spectra. Applying this measure on raw signal data is very sensitive, what leads to the fact that the presented method is not able to identify people walking together, even when they walk in-step.

In [5] the collective behavior of groups is studied. As the probands are not supposed to walk in-step, signals are not expected to be as related as in [4]. Wirz et al. resolve the issue by correlating basic features (mean and variance) of the signal. This is done by the use of windowed cross-correlation. Groups can be discriminated successfully because the individuals of a group have similarities in motion behavior, such as the speed and the direction of their motion.

In [6] wearable acceleration sensors are used to track similar gestures of ballet dancers. Who is leading or lagging in a ensemble of dancers and how they respond to each other with complementary movements is evaluated by windowed cross-correlation of the raw signals. Further an activity measure based on windowed variance is introduced and compared among the dancers to extract correlations of the activities observed.

Important to notice is that we do not expect to have similar signals for the task of classifying dancers and non-dancers. The mentioned projects suggest, that the correlation of *raw* signal data might not lead to a conclusion about the similarity of roughly related motion.

Other approaches for gesture tracking are presented in [7] and [8], where the movements are analyzed and mapped to sound to build a wearable music instrument. In [7] the data is recorded with a visual motion capture system. They work with two data sets, one for sound and the other one containing training data of the gestures. The training data is allocated by the use of Principal Component Analysis. For pattern recognition the similarity of gestures is analyzed by computing the Euclidean distance to the training set. The instrument presented by Takegawa et al. [8] uses acceleration sensors. To measure the similarity of two time-series Dynamic Time Warping is used, which allows recognition of the same sequence independently from the speed of an action. With an interface the dancers define the mapping of motion to sound before playing.

Recently the 'Sync-in-Team Game' [10] was presented, which is derived from 'The Musical Synchrotron'[9]. In [9] the challenge is to move the body synchronous to the beat of the played back music. With a visual feedback the dancers get an indication of how well they are synchronized. The goal is to stay with the beat of the music.

For the score they calculate the BPM[1] of the dancer. Prior a Fourier analysis the three directions of the windowed acceleration data is summed up and filtered. In the spectrum the location of the peak value is observed and its corresponding tempo in BPM is calculated. Afterwards the BPM value of the dancer is compared to the BPM value of the music and with a certain tolerance it is decided if the individual is dancing to the beat or not. For the 'Sync-in-Team Game' [10] the same algorithm was used to control a likewise play but further the influence of social interaction to the task was studied. This work relates audio and sensor signal features in an interesting manner, although the players do not dance arbitrarily but in a controlled loop of score and sound.

[1] Beats per Minute, a measure that is typically used to determine the tempo of a piece of music

Chapter 1

Experimental Setup

To derive methods for characterizing dancers and non-dancers we accomplished two different experiments with different configurations. In a lab-sized experiment characteristic data was collected to obtain a preliminary understanding about the acceleration signal of dancers and non-dancers, to evaluate different sensor locations and to find and test promising algorithms.

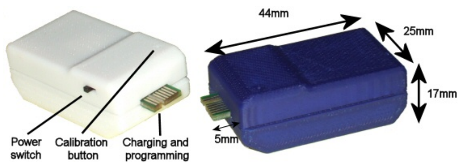

Figure 1.1: Acceleration sensors used for the experiments.

Thereafter a larger experiment has been carried out for the evaluation of methods that bear good prospects and for collecting data for related upcoming projects. In the second experiment we cared intensely to have a realistic setting and to have adequate data for interesting situations. The whole happening was documented with pictures and videos for both experiments. These recordings serve as a reference to keep unexpected events traceable. Incidents were documented manually.

1.1 Experiment I

For the first experiment the recording was done in a small lab room. We mainly focused on having specific settings and situations that may be critical to handle. We worked out 15 different configurations which held combinations from predefined, different expressions of motion:

- **Hand-Tapping:** Tapping with the hand to the beat of the played back music.
- **Toe-Tapping:** Tapping with the feet to the beat of the played back music. Either with the toes or with the heel, we cared to have both.

- **Free Dancing:** Dancing at will, no restrictions.
- **Free Moving:** Moving at will, alternating randomly between standing and walking, but not dancing.
- **Bouncing to Beat:** Bouncing to the beat of the played back music.
- **Walking to Beat:** Walking to the beat of the played back music.

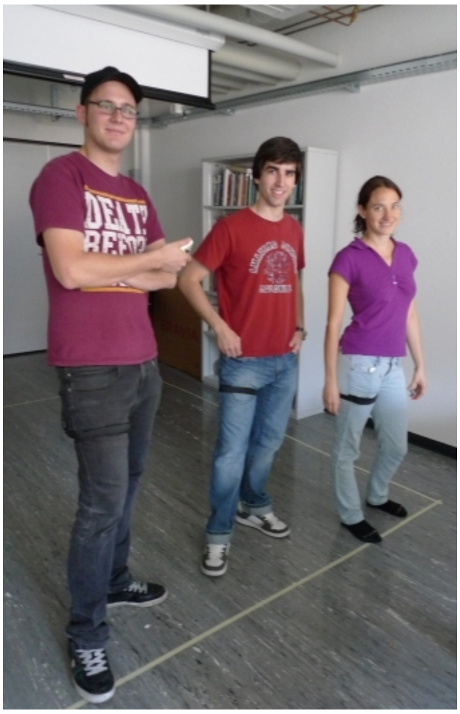

Figure 1.2: Probands of Experiment I with sensors held in hand and attached to leg.

As audio signals, three different songs from various music styles and a pure click signal were taken. The songs were chosen as representatives for the following three different setups:

- **Bar:** A song with a rather chilling groove, that typically is played at a bar or in a foyer of a club.
  (Song: Nouvelle Vague - Too Drunk to Fuck)
- **Straight:** A straight rock track, where the rhythm is clearly and easily perceptible.
  (Song: AC/DC - Rock'n'Roll Train)
- **Dance Floor:** A song that is played frequently at discotheques.
  (Song: Nelly Furtado - Maneater)

The categories ‘Bar’ and ‘Dance Floor’ were chosen to cover situations, where we believe that an automatic DJ system may be used in the future. To have a song that is easy to dance to also for not that gifted amateurs the section ‘Straight’ was added. The click signal served as an indicator to dispute the rhythmical sensitivity of the test people.

We had three probands, two male and one female, which were briefly instructed before the start of each configuration. For the selection a certain kind of heterogeneity was claimed, especially in the matter of dancing capabilities. The configurations can be found in Table 1.1. For configurations that deal with more than one kind of motion, the probands were partitioned into two groups.

Table 1.1: Experiment I: Configurations

Click	Bar	Straight	Dance Floor
Hand-Tapping	Hand-Tapping	Hand-Tapping	Hand-Tapping
None	Free Dancing vs. Free Moving	Free Dancing vs. Free Moving	Free Dancing vs. Free Moving
Walking to Beat vs. Free Walking	Free Moving vs. Walking to Beat	Bouncing to Beat vs. Free Moving	Free Moving vs. Walking to Beat
Toe-Tapping	Toe-Tapping	Free Dancing vs. Free Moving	Toe-Tapping

To capture data each proband was equipped with two 3-axis acceleration sensors, one held in hand and the other one worn on leg. We used two sensors per test person to elaborate the best positioning for this project. Additionally two participants were equipped with state-of-the-art mobile phones running an acceleration data recording application. The positioning was similar to the attached devices.

The sensors communicated via Bluetooth to a laptop by the use of the CRN-Toolbox [11].

The acceleration data was sampled at a rate of 64 Hz. The useful frequency range of human motion is said to lie between 0 Hz to 10 Hz [2]. The music was played back with a second laptop running Audacity[1]. We used wave audio files, mono track with a sampling rate of 44’100 Hz. To allow for synchronization of the sensor and audio data after the experiment, labels were set in real-time.

The experiment had a duration of one hour, each of the 15 configurations held 2 minutes of data, which corresponds to half an hour of observation data.

[1]A free, open source software for recording and editing sound
http://audacity.sourceforge.net

Figure 1.3: Probands of Experiment I, performing the following motions: (from left to right) dancing, dancing, free moving.

1.2 Experiment II

The second experiment was held in a realistic environment in a students bar near Berninaplatz in Zurich. We worked out seven scenes. We started with a set up and the last scene was completely unscripted. To cover more than a handful of music pieces, we decided to divide each scene into 2-minutes time slots. I.e. that a typical configuration holds parts of three different pieces of music and has a duration of 6 minutes. We had eleven participants, three women and eight men, with different educational background and dancing capabilities.

Figure 1.4: Sensor attached to leg.

To answer questions concerning the discrimination between dancers and non-dancers, the inspection of interaction in group dancing and dancing aborts, we worked out scenes including the following phenomena:

- Dancing vs. non-dancing
  - Non-dancers in a different location
  - Non-dancers in the same place as dancers
- Dance abandonment
- Groups of people dancing together

The script is shown in Table 1.2.

Table 1.2: Experiment II: Script

Scene	**Description**	**Duration**
0	Put up sensors, set up groups, serve drinks and snacks to participants.	15 min
1	Toe-Tapping with sensor-attached leg to click signal. Toe-Tapping with the other leg to click signal. Free Dancing	1 min 1 min 4 min
2	2 groups of 5 to 6 participants each: One group dances inside while the other one is asked to go outside and move freely. Interchange after 2 minutes	6min
3	Dancing in groups: 3 teams of 3 to 4 people are dancing faced to each member of the corresponding group. One group fills out a questionnaire in the first time slot.	6 min
4	While dancing three participants are chosen spontaneously. The selected people stop dancing to fill out a questionnaire. Hard abort during the last time slot with a previously settled sign.	6 min
5	Free Dancing in the first two slots. Polonaise in the last slot.	6 min
6	Unscripted part: The participants are free in their behavior. Possible options: dancing, standing at the bar, talking, etc.	30min

In Scenes 3 and 4 we tried to keep the selected probands in a natural, non-dancing motion by handing them out questionnaires. We chose this option to find out if a natural synchronisation to music occurs automatically. If so this would be fundamental for the task of discriminating dancers and non-dancers.

The data recording was done in a similar way as in the first experiment. We had to use two Bluetooth networks, because of the need for more sensors. Therefore we made more effort to keep the data adapted to synchronization. The data was sampled at 32 Hz for communication-stability reasons. One 3D acceleration sensor was attached at hip of each proband, three of them were supplementary equipped with mobile phones running an acceleration recording application. The phones were worn in pockets to get data similar to the attached sensors.

The second experiment took two hours and held approximately one hour of observation data.

Figure 1.5: Female probands of Experiment II, dancing. The strap of the leg attached sensor of the woman dancing in front is visible, she additionally wears a mobile phone in pocket, which is running an acceleration data recording application.

Figure 1.6: Probands of Experiment II, dancing.

Figure 1.7: Probands of Experiment II, group dancing to a polonaise.

Chapter 2

Signal Considerations and Algorithm Analysis

After the experiments the captured data was carefully analyzed. The analysis was done offline with MATLAB by the use of the standard Signal Processing Toolbox and the MIR Toolbox [12]. We focused on finding similarities between the sensor data and the played back audio. Therefore features from acceleration and audio data were extract by the use of existing methods. We looked for features that bear information that can be correlated in a way, that a decision can be made, whether the data observed corresponds to a dancer or not.

Figure 2.1 illustrates the anticipated approach to determine if people are dancing to music or not.

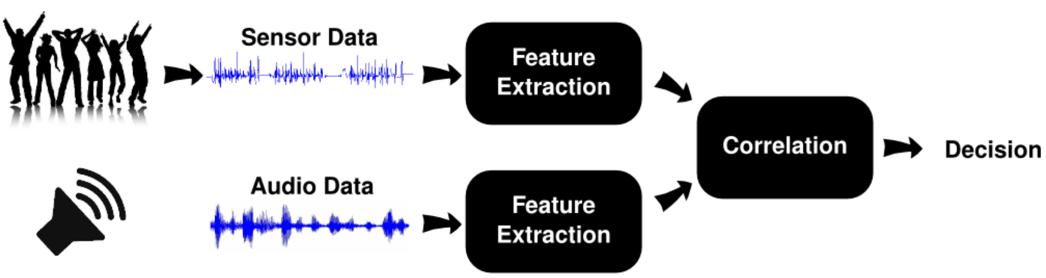

Figure 2.1: **Anticipated Approach:** Basic Features are extracted from audio and acceleration data. Afterwards these features are combined that a decision about motion behavior can be made. There are various methods for the task of extracting features from audio and acceleration data. We investigated on finding correlations of these features that allow us to determine if people are dancing to music or not.

2.1 Synchronization

For an in-time analysis of the recorded data the following three different data streams had to be synchronized:

- Sensors
- Networks
- Music

To allow synchronization of sensors, networks and music the data streams were marked in each recording with a bunch of labels. The labels were set manually. The possibility to average over labels kept the error insignificant due to human reaction rate. The synchronization among sensors in a given network is a feature of the CRN-Toolbox [11]. To keep the two Bluetooth networks in the same time line, the time stamp of the incoming packets had to be recomputed, which led to a maximal shift error of 1/32 seconds. Finally the synchronization of the music to the synchronized sensor networks was done.

2.2 Pre-Processing

We always processed the magnitude of the 3-axis outputs of the acceleration sensors to be insensitive to sensor orientation and therefore only relative measures could be taken into account for post-processing. The acceleration data was high pass filtered to get rid off the constant offset due to the gravitation of earth. The acceleration and the audio data was generally windowed by the use of a hamming window with an overlap of 50%.

2.3 Observed Signal Characteristics

There are two interesting limits that characterize the interaction of humans with music. On one side Winter [2] demonstrates that the useful frequency spectrum of human motion lies in the range of 0 to 10 Hz, on the other hand Noorden et al. [3] tell us that frequencies above 4 Hz are irrelevant to human rhythm perception.

We like to highlight the fact, that 4 Hz does not have to be the limit for music influenced motion, because a human is capable to dance or to move to certain song tempos at double time of the rhythm perceived and further this is no limit for harmonics that occur with certain expressions of motion. Harmonics are spectral components that are related to the fundamental frequency by integer multiples.

This theoretical background was taken into account when deciding about filtering.

By analyzing the sensor data of the first experiment in frequency domain, we observed the following characteristics concerning sensor positioning:

- Data caught from leg attached sensors show a clear maximum peak at the frequency corresponding to the main, half or double tempo of the played back song.
  We concluded that a dancer mainly synchronizes to the beat with leg motion.
- Data caught from held in hand sensors *may* show a peak at the frequency corresponding to the main, half or double tempo of the played back song.
  We concluded that arm motion balances the dancing motion, varies the expression and is especially used for defining dancing style and for communicating to other dancers.

The observations are illustrated in Figure 2.2.

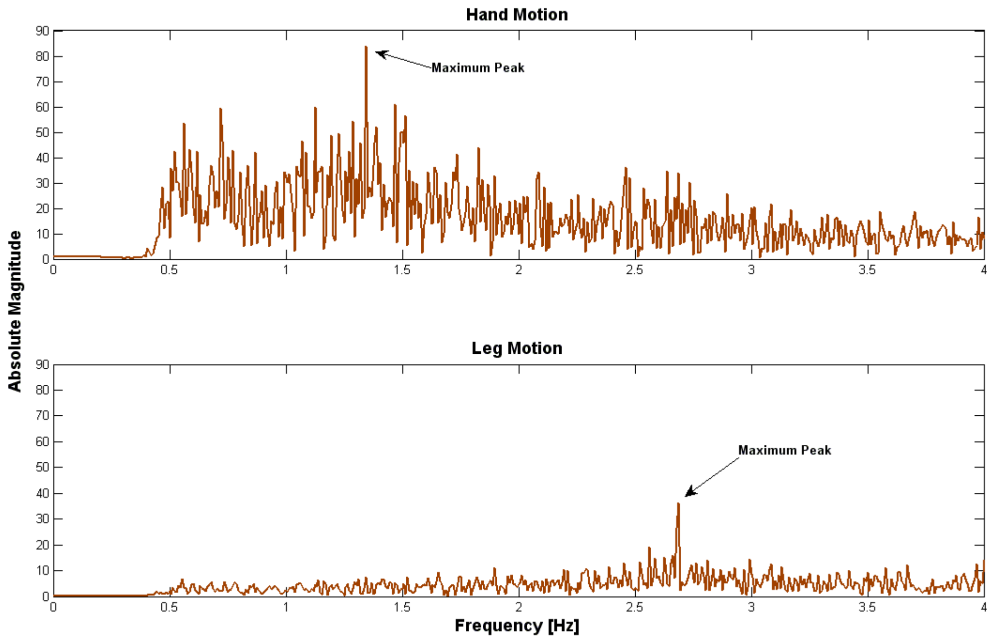

Figure 2.2: Spectra of hand and leg motion of a dancer. The maximum peak observed in leg motion is more dominant than the maximum peak observed in hand motion. The magnitude of the hand motion is higher because this sensor is placed at an extremity. The music the dancer is dancing to has a tempo of 160 BPM, roughly 2.7 Hz that is. The location of the maximum peak in leg motion corresponds to the main tempo of the song. The location of the maximum peak in hand motion corresponds to the half tempo of the song, 80 BPM, roughly 1.4 Hz that is.

For the mentioned reasons and the fact that people often wear their mobile phones in pockets, we decided to only focus on the acceleration signal of sensors worn at leg for the second experiment.

We further observed that expressive dancers show a higher spread of the maximum peak in frequency domain compared to dancers that insist mainly on the rhythm. Toe-, hand-tapping, walking and bouncing to the beat typically show narrow peaks in spectra and often feature several harmonics. The most numerous harmonics occurred with toe-tapping, an observed spectrum is shown in Figure 2.3.

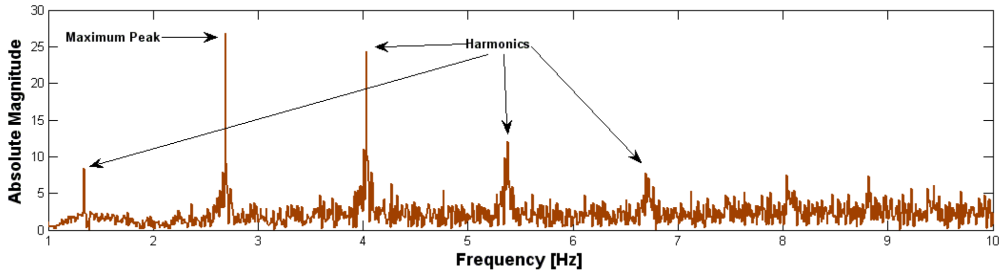

Figure 2.3: Toe-tapping and whipping motions typically show a lot of harmonics in spectrum of data caught by the leg sensor. Here the spectrum of the signal captured from a toe-tapping person is plotted. The acceleration data was not filtered. The tempo of the played back audio track is 160 BPM, roughly 2.7 Hz that is. The dominating peaks are approximately located at 1.4 Hz (half tempo), 2.7 Hz (main tempo), 4.1 Hz (3/2 tempo) and 5.4 Hz (double tempo).

The frequency responses of people that are standing around and moving just slightly, have a ‘low pass’ characteristic. An observed example can be found in Figure 2.4.

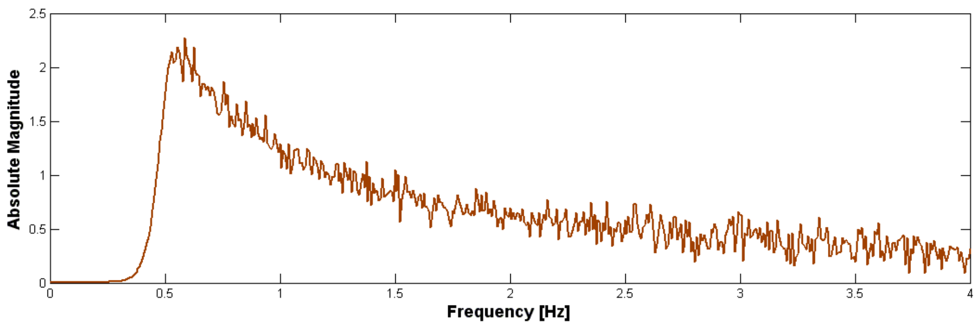

Figure 2.4: Data caught from a person standing around and moving slightly typically shows a ‘low pass’ characteristic in spectrum of data caught by the leg sensor.

Synchronizing to the music when dancing tends to be a matter of dancing style, although our probands mostly synchronized automatically to the beat, without actually asking them to do that. That dancers generally move to the rhythm and express themselves *dependent* of the played back music is an important fact, because dancing independently from music would not allow us to decide if someone is dancing by combining basic features from acceleration and audio data.

Figure 2.5 illustrates that acceleration signals caught from dancers and the corresponding audio signals generally do not seem to have evident similarities.

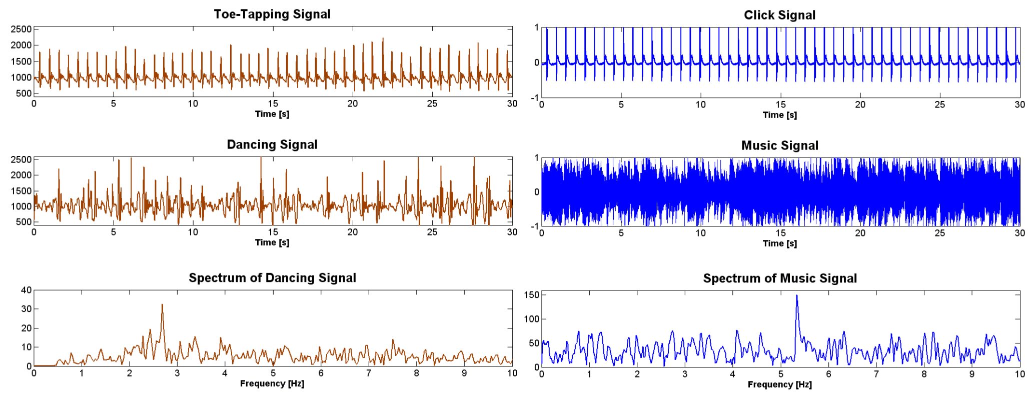

Figure 2.5: **Acceleration signals and corresponding audio signals in time and frequency domain:** In the artificial set up of toe-tapping to a click signal, signals look very similar in time domain. If we compare signals caught from a realistic set up, such as dancing to music, the similarity of time domain signals seems to get lost. If we transform these signals into the frequency domain, we obtain two peaks that are related to the tempo of the played back music. The tempo of the music signal is 135 BPM, which corresponds approximately to 2.7 Hz.

In the spectra of Figure 2.5 we obtain peaks, that are related to the tempo of the corresponding signal. In music BPM[1] is a widely used measure to define the tempo of a piece of music. The observed properties afford defining the tempo of a signal that is captured from leg-attached acceleration sensors:
The location of the maximum peak observed in the spectrum of such an acceleration signal corresponds to the main tempo of the performed dancing motion.

[1] Beats per Minute, a measure that is typically used to determine the tempo of a piece of music. If a song is said to have a main tempo of 120 BPM the double tempo corresponds to 240 BPM and the half tempo corresponds to 60 BPM. 120 BPM correspond to a motion frequency of 2 Hz.

2.4 BPM Algorithm

Based on the observed signal properties we implemented an adapted version of the algorithm presented in [9] for the task of discriminating dancers and non-dancers. We refer to it as the BPM Algorithm. The signal flow diagram is shown in Figure 2.6.

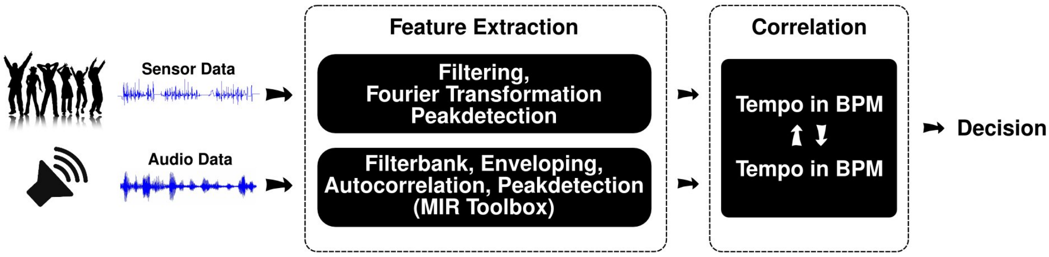

Figure 2.6: **Signal flow diagram of the BPM Algorithm.** With separate processing streams for audio and acceleration data the tempo of the acceleration data and audio data is computed. In a correlation step the tempi are compared and afterwards a decision is made if the person corresponding to the processed acceleration data is dancing or not.

The algorithm works as follows:

1. Low pass filter the acceleration data with an accurate cut-off frequency[2].
2. Calculate the spectrum of the acceleration data by applying a FFT.
3. Look for the location of the maximum peak observed in spectrum and compute the corresponding BPM value.
4. Compare the found BPM value to the main, half and double tempo of the played back music.

The decision is made with a certain relative tolerance. We defined the tolerance in values of ±BPM. It determines the allowed shift between the BPM of a dancer and the BPM of the played back music. The tolerance is relative in the sense that in case of double tempo, we allow twice the tolerance and in case of half tempo, we allow half the tolerance.

The tempo of the played back music was computed for each window by the use of the *mirtempo()* function of the MIR Toolbox[12]. The audio data is first filtered with an array of band pass filters. From the filterbank the energy envelopes are calculated and auto-correlated. With a peak detection mechanism the tempo is extracted.

[2] The double tempo of the played back music should be detectable in the spectrum of the acceleration signals. We used 5 Hz as cut-off frequency for the evaluation of the algorithm, because the maximum double tempo observed in the music signals of the data set corresponds to 4.5 Hz

The computing time needed to extract the tempi of an acceleration data segment and the corresponding audio signal segment is in the order of seconds.

2.4.1 Algorithm Analysis

The evaluation was done on the data recorded in Scene 2 of Experiment II, cf. Table 1.2. In this scene we simultaneously have dancers and non-dancers. Further they interchange parts about half of the scene. To avoid side effects we extracted two sections holding 230 seconds totally. Each section holds data from 10 acceleration sensors and the two sections feature a different audio track. This allows us to evaluate the algorithm with 10 different people, each acting as dancer and non-dancer. It is to say that a large packet loss could be observed in Scene 2, see Section 2.5.

The lost segments were not taken into account for decision.

In Figure 2.7 the algorithm performance is plotted for various window sizes and tolerances. The best performance was observed with a window size of 25 seconds and a tolerance of ±3 BPM. The Accuracy[3] Curve for a window size of 25 seconds is displayed in Figure 2.8. In Figure 2.9 the decision history by time slots is illustrated.

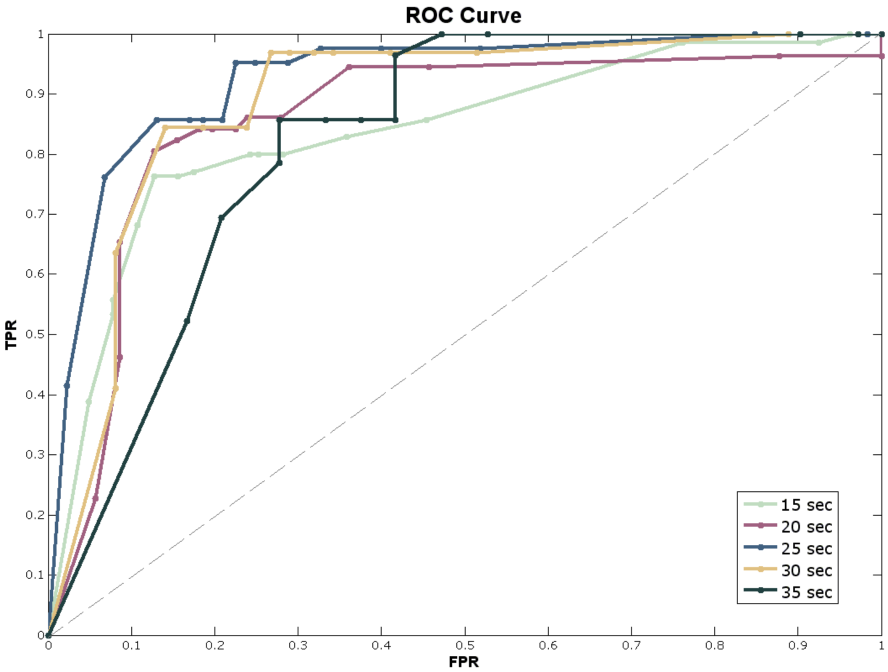

Figure 2.7: ROC Curve of the BPM Algorithm by Window Size. The True Positive Rate (TPR) counts the amount of dancers classified as dancers. The False Positive Rate (FPR) counts the amount of non-dancers classified as dancers.

[3] Amount of correct decisions divided by all decisions.

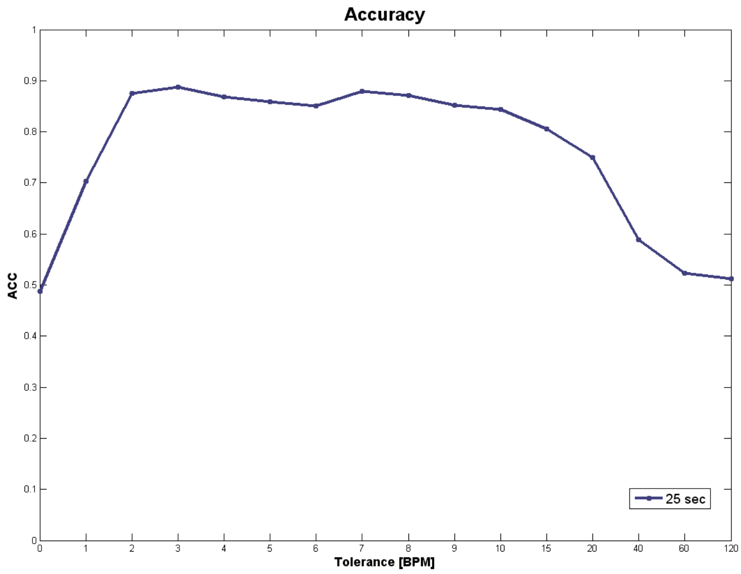

Figure 2.8: Accuracy Curve of the BPM Algorithm by Tolerances. Data windowed by a hamming window of 25 seconds with 50% overlap. The maximum accuracy of 89 % is achieved with a tolerance of ±3 BPM.

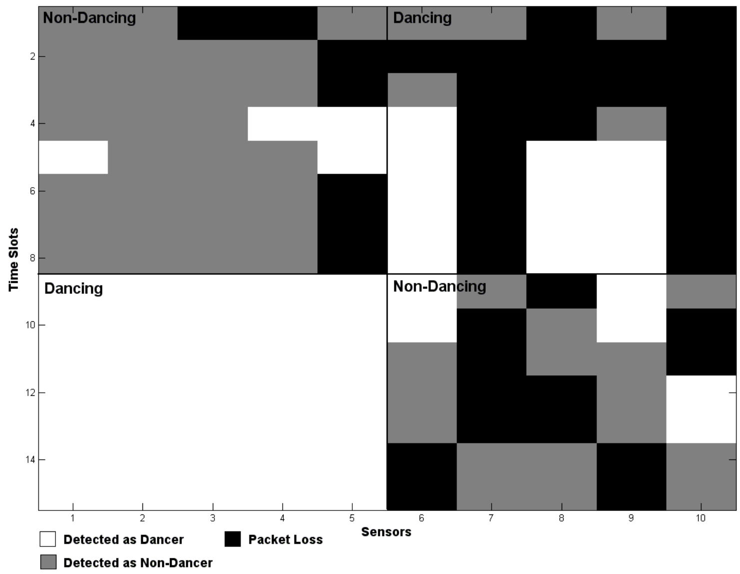

Figure 2.9: Decision History by Sensor and Time Slot of the BPM Algorithm. Data windowed by a hamming window of 25 seconds with 50% overlap. Decisions are made with a tolerance of ±3 BPM

Further a section from Scene 3 (cf. Table 1.2) was extracted to analyze the decisions made. This section holds the data where one group fills out a questionnaire and the other two groups are dancing. The section is 30 seconds long. The algorithm classifies almost all test subjects correctly, one non-dancer is misclassified because of the peak obtained at the double tempo of the played back music. In Figures 2.10 and 2.11 the spectra of the acceleration data of this section are shown.

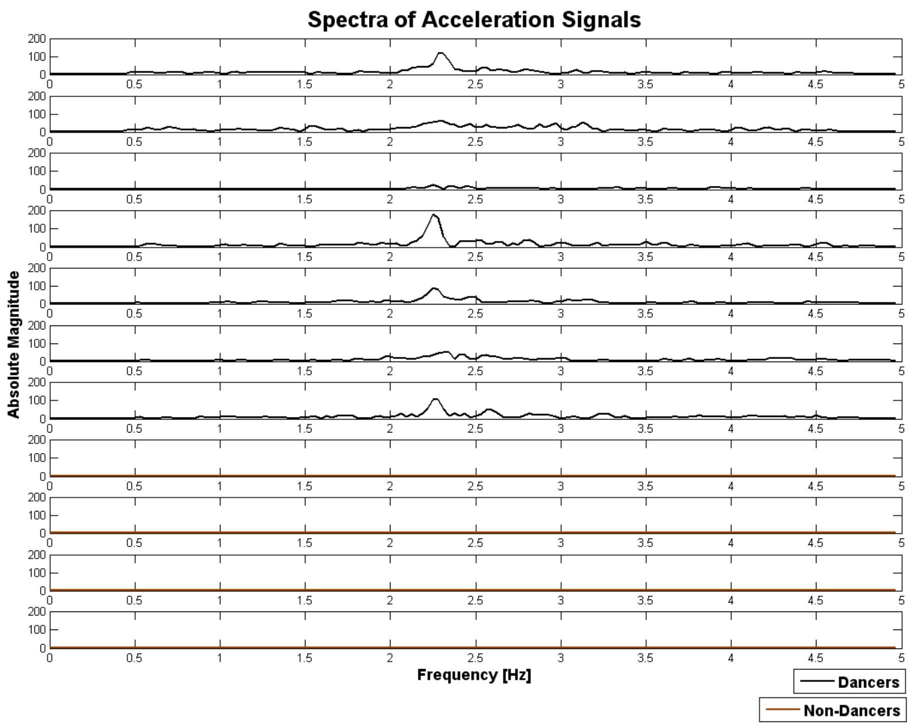

Figure 2.10: Spectrum of sensor data caught from dancers (1-7) and non-dancers (8-11). The tempo of the played back song is 135 BPM, which corresponds to 2.25 Hz. All dancers are classified as dancers by the algorithm. One non-dancer is declared as dancer, compare Figure 2.11.

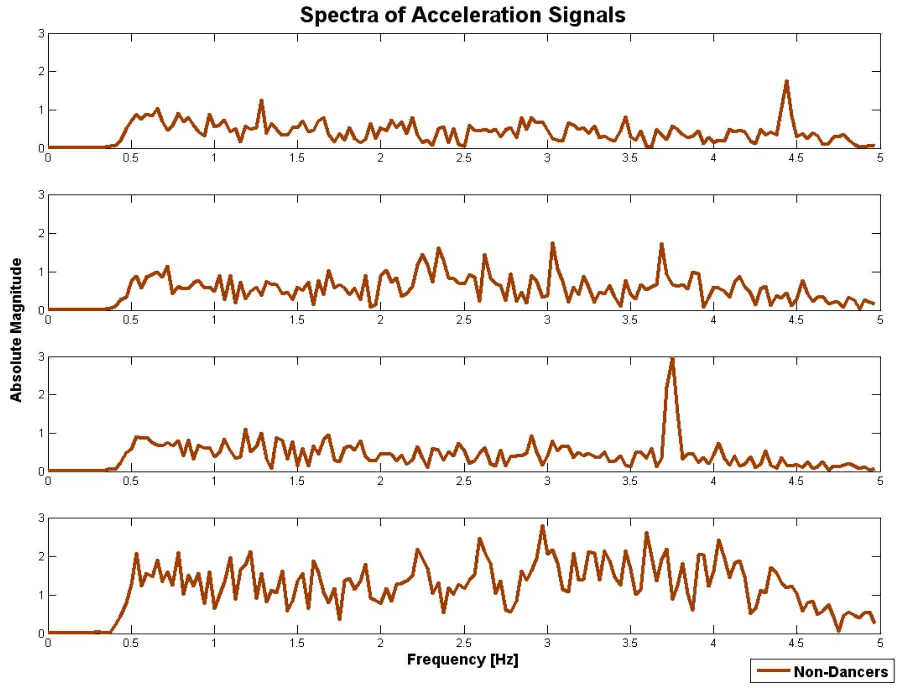

Figure 2.11: Detailed plot of non-dancers filling out a questionnaire. The first non-dancer is declared as dancer by the algorithm, because the location of the maximum peak corresponds to the double tempo of the played back song, 270 BPM, 4.5 Hz that is. The other three non-dancers are correctly classified as non-dancers.

2.4.2 BPM Algorithm Modifications

We made two modifications on the BPM Algorithm and evaluated them on the same data set:

- We took the locations of the second and the third maximum peak value in the frequency spectra of the acceleration signal into account for decision. If one of the locations corresponds to the main, half or double tempo of the played back music, the corresponding person is classified as dancer.

- In a second evaluation, we allowed comparing the tempo observed in acceleration signals to the *main tempo* of the music only.

The performance of the algorithm considering the three maximum peaks is displayed in Figure 2.12. The best performance was observed with a window size of 10 seconds and a tolerance of ±6 BPM. In Figure 2.13 the Accuracy Curve for a window size of 10 seconds is shown. The second modification was not able to beat the observed performance of the first modification.

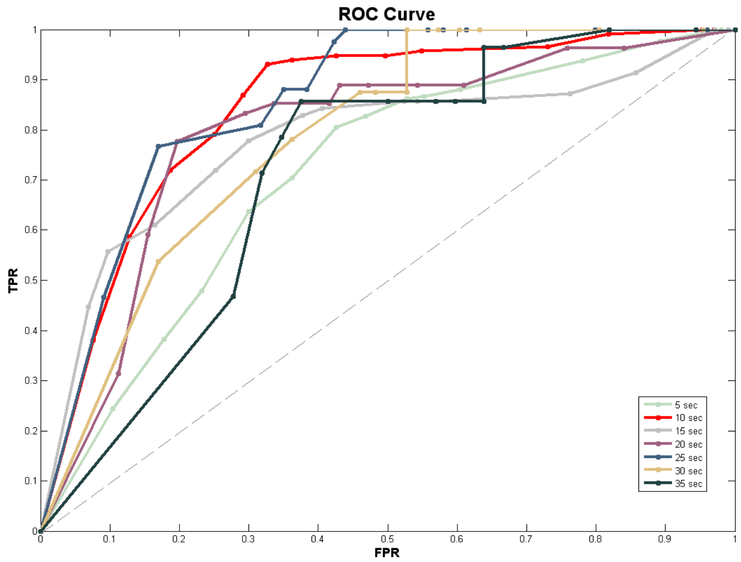

Figure 2.12: ROC Curve of the Modified BPM Algorithm by Window Size. The True Positive Rate (TPR) counts the amount of dancers classified as dancers. The False Positive Rate (FPR) counts the amount of non-dancers classified as dancers.

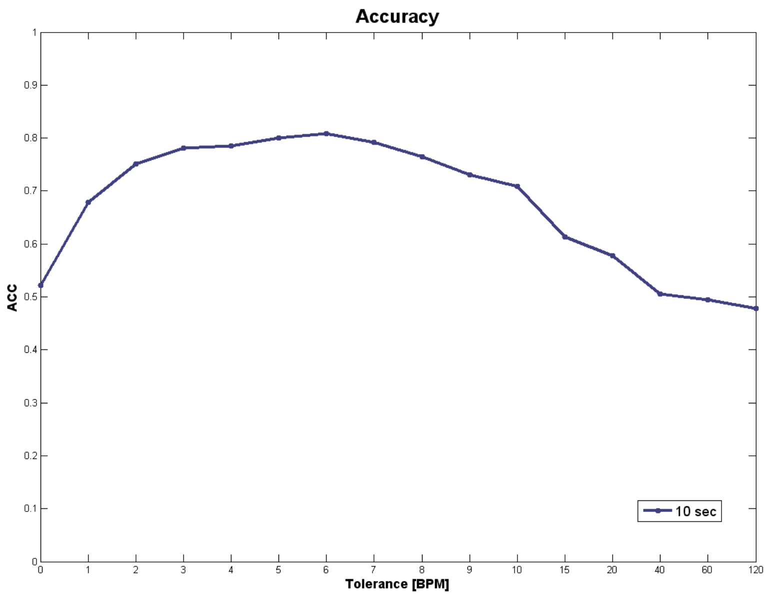

Figure 2.13: Accuracy Curve of the Modified BPM Algorithm by Tolerances. Data windowed by a hamming window of 10 seconds with 50% overlap. The maximum accuracy of 81 % is achieved with a tolerance of ±6 BPM.

2.4.3 Other Algorithms

We further tried approaches inspired from audio and image processing research.

We implemented a Census Transformation [13], where the time signal was windowed and divided into three segments. Relative to the energy or variance a bit value was assigned to neighbor segments. Similarity was expressed with the Hamming Distance.We tried various window sizes, but the acceleration signals did not correlate with the audio signals.

Finally, we tried an approach that compares the first three moments[4] [14] of the filtered audio and acceleration data in a range of 0 Hz to 5 Hz, but the spectra generally seem not to show common features.

2.5 Discussion

The presented methods are capable to discriminate dancers and non-dancers, but are not able to discriminate people that are dancing or just moving *to the beat*. We observed that the BPM Algorithm shows the best performance for a window size of 25 seconds on the evaluation data set. A certain window length is needed to track dancers as dancers, because of their individual motion behavior. When window size reaches a given limit the algorithm performance starts to decrease. This is due to the fact that the algorithm is more likely to detect non-dancers as dancers for larger window sizes.

The modifications we made to the BPM Algorithm allowed us to get an accurate classification for smaller window sizes. But as may be expected the performance decreases faster for larger windows, because the algorithm is more sensitive in detecting dancers.

It seems that tolerance has a dependence on window size, therefore we prospect to decide on tolerance and window size simultaneously for a given application.

We think that the presented method is an interesting feedback for an automatic DJ application, especially because of the moderate window sizes and the rather low computing time consumptions.

The wrong classification of the non-dancer in Scene 3 (cf. Figure 2.11), was analyzed by watching the recorded video data. It may be that the proband moved slightly with the beat, what could make sense due to the harmonics observed, but the legs are not visible in the video and therefore this is just guessing.

As can be seen in Figure 2.9 we had a lot of communication errors in Scene 2 (cf. 1.2). The errors occurred because of the location change of the groups, although we accompanied them with the laptop the sensors communicated with.

Finally it remains to say that we observed a small constant tempo drift compared to the acceleration data, which has been analyzed and has been assigned to the playback tempo related to the software and/or hardware used. For this reason we

[4] Spectral Centroid, Spread and Skewness of the Spectral Distribution.

decided to record the audio signal in the second experiment with a microphone for completeness.

Conclusion

Conclusion

Generally it is to say that audio data and data caught from dancers wearing one acceleration sensor at leg seem not to have a lot of strong similarities, but we observed that dancers typically synchronize to the tempo of the music they are dancing to. This allowed us to extract the tempo as common feature of the two data streams. We presented a method that uses this similarity for characterizing dancers and non-dancers. The implemented algorithm is capable to decide accuratly if someone is dancing or not, which is an interesting information for a feedback for an automatic DJ system. On the evaluated data set the algorithm performed best for a window size of 25 seconds. We further made modifications to the algorithm, which were not able to break the performance of the standard implementation for data window lengths of 20 to 30 seconds. But the modified algorithm showed a better performance for data segments on a 5 to 10 seconds basis. Although all presented methods are not able to discriminate people that are dancing or moving *to the beat*. But for the use within an automatic DJ system this may be irrelevant, because we think that motions like moving or whipping *to the beat* are also expressions of liking the played back music.

Future Work

Combining features among sensor signals only could lead to an enhancement of the detection rate. Especially comparing the first three moments of spectral distribution among sensor signals could be worth investigating. This may allow to characterize the performed movements more exactly, such that a more precise decision could be made. Beside investigating similarities of sensor signals the interaction and the communication of dancers, especially in group dancing could lead to a better understanding of how people express their feelings with dancing.

To allow the methods to be used in a real-life setting the data caught by the mobile phones should be investigated and analyzed.

Acknowledgements

We would like to thank the people of the ‘Netzwerk Bar’ for providing the location for the second experiment and all the test subjects for their impressive dancing commitment, their patience and their will to participate in our experiments.

Bibliography

[1] J. Adshead-Lansdale and J. Layson. *Dance history: An introduction.* Routledge, 1994.

[2] D.A. Winter. *Biomechanics and motor control of human movement.* Wiley, 2009.

[3] L. Van Noorden and D. Moelants. Resonance in the perception of musical pulse. *Journal of New Music Research*, 28(1):43–66, 1999.

[4] J. Lester, B. Hannaford, and G. Borriello. "Are You with Me?" Using accelerometers to determine if two devices are carried by the same person. In *Proceedings of Second International Conference on Pervasive Computing (Pervasive 2004*, pages 33–50. Citeseer, 2004.

[5] Martin Wirz, Daniel Roggen, and Gerhard Tröster. Decentralized Detection of Group Formations from Wearable Acceleration Sensors. In *CSE (4)*, pages 952–959, 2009.

[6] Ryan Aylward and Joseph A. Paradiso. Sensemble: A Wireless, Compact, Multi-User Sensor System for Interactive Dance. In *NIME*, pages 134–139, 2006.

[7] F. Bevilacqua, J. Ridenour, and D.J. Cuccia. 3D motion capture data: motion analysis and mapping to music. In *Proceedings of the Workshop/Symposium on Sensing and Input for Media-centric Systems.* Citeseer, 2002.

[8] Y. Takegawa T. Terada M. Fujimoto, N. Fujita and M. Tsukamoto. "A Motion Recognition Method for a Wearable Dancing Musical Instrument". In *Proceedings of the 2009 International Symposium on Wearable Computers.*

[9] M. Demey, M. Leman, F. Bossuyt, and J. Vanfleteren. The Musical Synchrotron: using wireless motion sensors to study how social interaction affects synchronization with musical tempo. In *Proceedings of the 8th International Conference on New Interfaces for Musical Expression*, 2008.

[10] Marc Leman, Michiel Demey, Micheline Lesaffre, Leon van Noorden, and Dirk Moelants. Concepts, Technology, and Assessment of the Social Music Game "Sync-in-Team'. In *CSE (4)*, pages 837–842, 2009.

[11] D. Bannach, K. Kunze, P. Lukowicz, and O. Amft. Distributed modular toolbox for multi-modal context recognition. *Lecture Notes in Computer Science*, 3894:99, 2006.

[12] O. Lartillot and P. Toiviainen. MIR in Matlab (II): A toolbox for musical feature extraction from audio. In *International Conference on Music Information Retrieval*, pages 237–244, 2007.

[13] K. Yamada, K. Mochizuki, K. Aizawa, and T. Saito. Motion segmentation with census transform. *Lecture Notes in Computer Science*, pages 903–908, 2001.

[14] G. Peeters. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. *CUIDADO IST Project Report*, pages 1–25, 2004.