Sensing Dance Engagement for Collaborative Music Control

Michael Kuhn, Roger Wattenhofer
Distributed Computing Group, ETH Zürich
{kuhnmi, wattenhofer}@tik.ee.ethz.ch

Martin Wirz, Matthias Flückiger, Gerhard Tröster
Wearable Computing Lab., ETH Zürich
{martin.wirz, maflue, troester}@ife.ethz.ch

Abstract

We introduce a concept that allows attendees of a party to collaboratively influence the music selection process. As explicit feedback is likely to disturb the atmosphere, we introduce an unobtrusive, implicit feedback mechanism. In particular, we propose to sense the partygoers' dance engagement by means of their mobile phones. Since people tend to dance more when they enjoy the music, this metric provides a measure of the crowd's satisfaction. We verify experimentally that music can be selected such that the generated song sequence matches the taste of the crowd, and we demonstrate the feasibility of inferring an individual's dance engagement using the mobile phone's accelerometer. Our experiments show that the proposed method is robust to the body location at which the phone is worn and reaches an overall accuracy of 86.8% *for distinguishing dancers from non-dancers.*

1 Introduction

Music plays an integral role when people socialize. Often, music is not just passively consumed, but people actively engage in dancing at parties or in clubs. We believe that in such settings a combination of smartphone based sensing with novel approaches to music information retrieval could add to the traditional, DJ controlled, music selection process by making it more interactive and thus enjoyable.

In this work we provide evidence that dancing can be seen as a sign of agreement to the music. We propose to exploit this phenomenon to steer song selection such that the crowd's acceptance of the music improves over time. For this purpose, we show that concepts previously used to generate playlists for individuals apply to the group case when participants can provide explicit feedback. To overcome the necessity of explicit voting, we then propose to exploit the smartphones of the attendees of a party to measure their dancing behavior. Hereby, a single dancer basically provides an implicit "vote" for the played song, and the cumulated "votes" give a picture of the mood of the crowd that can be used to steer song selection according to the crowd's needs.

We conducted experiments that demonstrate the feasibility of this approach. In particular, we show that we can reliably discriminate dancing from other activities that commonly occur in dance clubs or at parties by means of the smartphones' acceleration sensors. Rather than classifying dancers solely from the acceleration data, we take advantage of the correlation between the acceleration and the audio signal.

2 Related Work and User Survey

Several authors have identified the need for audience dependent music selection in public settings and propose to use explicit feedback to control the music selection [7, 10, 11]. We believe that in a party setting users feel distracted when having to provide explicit feedback. Thus, we propose to improve upon the mentioned approaches by introducing a sensor controlled implicit feedback channel. Such sensor-based feedback mechanisms have been used mainly to generate audio-visual effects in interactive environments. A wide majority makes use of wearable accelerometers [1, 4]. Another type of interactive music system using wearable sensors are social games, such as the Musical Synchrotron [5]. These games show that the tempo of a user's motion can accurately be measured using accelerometers and compared to the tempo extracted from a music piece.

Anecdotal evidence suggests that people tend to dance more when they like the music, and that dancing and mood are also positively correlated. Moreover, the scarce literature on music perception and its influence on dance behavior and mood suggests a similar coupling. A positive affect to infants could be shown when they successfully synchronize to music [17], and it is known that pleasurable music and rhythmic movement can simulate reward areas in the brain [8]. It is further known that there is a relationship between personal rhythm preference and the activation of the brain's motor system [6]. These results might explain why people go to nightclubs and enjoy dancing, but they are not general enough to be conclusive. To get a broader view we have conducted an online survey. This survey was filled in by 234 subjects (69% male, 41% female) from western Europe with an average age of $\mu = 24.75$ years (min: 16, max: 59, stdev. σ: 5). First, a number of factors had to be judged on a 5-point Likert scale to understand how strongly different factors influence dancing. We found that the favorite music is the most important factor (4.44) followed by the general party atmosphere (4.21) and famous songs (4.11). Other factors, such as the DJ (3.65) and alcohol (3.29), were said to have a smaller influence. A majority agreed that dancing is "important" or "very important" for a good atmosphere.

These results indicate a positive effect of dancing to the mood and of pleasant music to dancing. Moreover, they suggest that there is a feedback loop: people tend to dance more if the general mood is good, and in turn, people's mood improves when they are dancing. Moreover, this feedback loop can be stimulated by selecting appropriate music.

3 Steering Music by Explicit Feedback

To foster the crowd's engagement, we propose to create a playlist dynamically, controlled by the feedback gathered from the audience. For this purpose we use a variant of the algorithm proposed by Bossard et al. [3] which deals with dynamic playlist generation for single users using skipping behavior as feedback to identify regions in music similarity spaces that match the user's mood. Rather than looking at the user's skipping behavior, we assume that each user can cast a vote in a thumbs-up/thumbs-down style for each song. We then convert the cumulated votes for a given song into a single rating. As an underlying music similarity space we rely on the recently proposed social audio features [9].

For evaluation we conducted an experiment with approximately 50 students in a small bar. 17 of the party guests were equipped with smartphones that could be used for voting. The duration of the experiment was approximately 90 minutes and comprised a total of 25 songs from a collection of roughly $2,500$ songs.

The ratings corresponding to the user feedback were not only used to control the algorithm, but also to estimate the satisfaction of the audience. To get a ground truth, we interspersed some random songs into the generated song sequence. We found that the ratings of songs chosen by the algorithm (median: 0.46, μ: 0.46, σ: 0.24) were considerably higher than those of random songs (median: 0.3, μ: 0.29, σ: 0.19). A ranksum test rejects the null hypothesis of equal medians at the $\alpha = 0.2$ level. These values indicate that the algorithm was able to find regions that match the taste of a majority of party-goers, but data is too sparse to be conclusive.

4 Implicit Feedback through Accelerometers

As argued in Section 2, dancing can be seen as an implicit, real-time feedback serving as a vote. The higher the fraction of people engaging with the music, the more we expect them to approve it.

In this section, we present a method to discriminate dancers from non-dancers using wearable acceleration sensors. Spontaneous engagement can also come in less intense ways than dancing, such as in form of toe-tapping – a rhythmic lifting of the toes or the heel of one foot. We want to be able to discriminate the two behaviors. Moreover, it is important to detect other activities commonly performed during parties, namely walking (e.g. when going to the bar to order a drink), and standing around, such as when having a chat. These considerations result in a classification problem with the four classes *dancing*, *toe-tapping*, *walking*, and *standing around*.

Acceleration sensors have been used intensively in the wearable computing community to discriminate activities [2, 12]. The ubiquitous availability of these sensors in mobile phones motivates our use to infer the dancing state of clubbers. Before describing how we distinguish dancers from non-dancers, we briefly discuss some relevant phenomena of the way humans synchronize their movements to music.

4.1 Synchronizing to Music

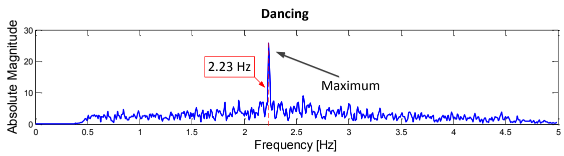
From existing literature we know that people naturally synchronize their movements to music by following the perceived tempo (measured in beats per minute, BPM) [13, 5].

Two factors limit the interaction of humans with music. On the one hand, Winter [15] 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. [14] state that frequencies above 4 Hz are irrelevant to human rhythm perception. Moreover, Styns et al. [13] show that when moving to fast music, people sometimes synchronize to half the tempo. The opposite is the case when moving to slow music, i.e. people sometimes synchronize to twice the tempo.

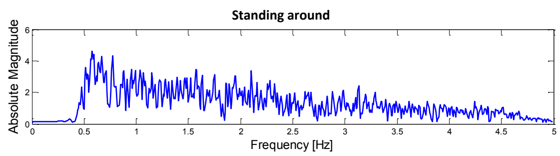
We expect to see a distinct peak in the frequency spectrum of the resulting acceleration signal, corresponding to the tempo of the movement. In case of motion that is in sync with the music, the location of this peak is expected match the tempo of the music (or twice or half this tempo as discussed before). These phenomena are illustrated in Figure 1.

4.2 Algorithm

As the dance detection has to work in real-time, we chose a window-based approach. Over this window, we calculate a feature vector $\vec{f}$ consisting of two features: the signal energy E_{Acc} and the tempo deviation ΔBPM. While we expect the energy to help distinguish between low (standing



(a) Spectrum of the acceleration signal captured from a dancing person. The tempo of the played back audio track is 133 BPM ($= 2.22$ Hz). The dominant peak is located at 2.23 Hz corresponding to the main tempo.



(b) Acceleration data from a person standing around or moving in place typically shows a low pass characteristic in the spectrum.

Figure 1. Magnitude of the spectrum of different acceleration signals.

around, and toe-tapping) and high (dancing, walking) intensity movements, the tempo deviation is supposed to discriminate activities that are synchronized with music (toe-tapping, dancing) from non-synchronized activities (walking, standing around).

The signal energy E_{Acc} of the acceleration signal is calculated in the time domain. ΔBPM measures the minimal difference between the tempo of the music the location of the "relevant" spectral peak in the motion signal. Recall that people sometimes synchronize to half or twice the tempo of the music; thus, ΔBPM is given as the minimal difference to any of these locations.

The tempo of the acceleration signal was calculated using the following algorithm:

1. Low pass filter the acceleration data with a cut-off frequency of 5 Hz ($= 300$ BPM). This cut-off is well in the limits imposed by the previously discussed phenomena in the context of human synchronization to music.

2. Calculate the spectrum of the acceleration data by applying a Fast Fourier Transform (FFT).

3. Detect the location of the maximum peak observed in the spectrum and compute the corresponding BPM value.

The music tempo was computed by using the MIRTempo function of the MIRToolbox Matlab library.

4.3 Experiment

To verify our approach, we designed and conducted a lab experiment. We equipped two female and two male subjects with Android devices. To investigate the influence of different sensor placements each subject wore a mobile phone in the trousers' front and back pocket and in one hand. Female subjects had an additional phone in their hand bag. This arrangement covers the placements where phones are carried more than 90% of the time [16].

The experiment followed a predefined script. We played 2min excerpts of three songs of different genres and tempi. Each song was played twice. During the first playback, the subjects were asked to dance. During the second playback, they were asked to toe-tap. At the end of all sessions, snacks were served to the participants causing them to stand around naturally. Additionally, each subject had to walk around the floor of the university building. A total of 35min of acceleration data was recorded for each subject with a sampling rate of 32 Hz (as given by the mobile phones).

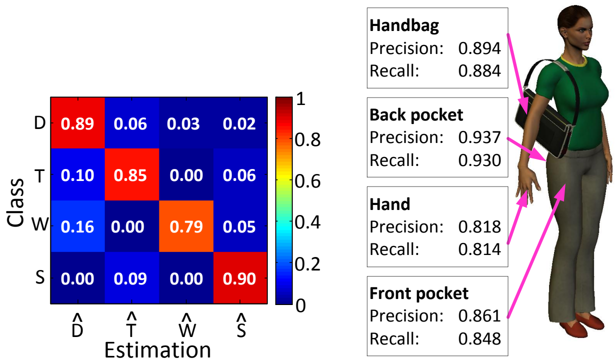
4.4 Analysis

Two main aspects were analyzed: (1) The discrimination of the dancing activity from other activities commonly performed at parties, and (2) the influence of the body locations at which the sensors are worn.

In our lab experiment, we considered the four classes *dancing*, *toe-tapping*, *standing around*, and *walking around*. To investigate the performance of our algorithm in distinguishing these four different activities, we classified the feature vector $\vec{f}$, as given by E_{Acc} and ΔBPM. For evaluation, we used a window length of 25s. The cumulated sensor information from all mobile phones and all individuals was used for training, i.e. training was performed subject and location independent.

Naïve Bayes classification achieved a precision of 0.88 and a recall of 0.87 for the four class problem as evaluated using a 10-fold cross-validation strategy. Figure 2(a) shows the corresponding confusion matrix. We were further interested in just discriminating dancers from non-dancers (*walking*, *standing around*, and *toe-tapping*). For this two class problem, we achieved a precision of 0.93 with a recall of 0.93. Another two-class problem distinguishes between engaging (*dancing* and *toe-tapping*) and non-engaging (*walking* and *standing around*) activities. For this problem, a precision of 0.91 and a recall of 0.93 was reached.

In real settings, we can not dictate where the phone is worn. Hence, our algorithm has to be robust with respect to different body locations. As seen in the previous section, good results can be achieved without considering any location influence. To better understand the influence of different locations, we applied the classification model of the 4-class problem to the feature vectors of each location independently (without retraining). Figure 2(b) summarizes the obtained results.



(a) Confusion matrix of the 4-class problem with the classes *dancing* (D), *toe-tapping* (T), *walking* (W) and *standing around* (S) and their estimations

(b) Classification accuracy for different sensor placements. The highest accuracy was achieved with sensors in the back pocket.

Figure 2. Classification results

4.5 Real World Verification

To evaluate our approach in a more realistic scenario, 11 subjects (8 male, 3 female, between 22 and 29 years old) participated in the experiment which was conducted in a small bar. To ensure a natural atmosphere, we played ambient music and offered drinks and snacks to the subjects before the experiment started. During this time, the subjects were instructed about the recorded signals. Music was played according to a predefined playlist containing two minute excerpts of different songs. During the experiment, we split the subjects into two groups. While one group was dancing to the music, the other group was outside, freely moving around (*walking*, *standing around*). Afterwards, the roles were switched. Our algorithm could discriminate *dancing* from the other activities with an overall accuracy of 0.89, which is comparable to the performance found in the lab experiment.

5 Conclusion

In this work we have investigated novel approaches that address the selection of music during parties. In particular, we have studied the feasibility in selecting music such that the resulting song sequences match the taste of a large fraction of participants. We proposed to use the level of people's dance engagement as a natural measure for their satisfaction with the played music. Taking this signal as an input, the envisioned system can automatically detect the style of music provoking maximal engagement of the crowd. On the one hand, it could thereby prove valuable to support a DJ. In settings that lack a human DJ, on the other hand, a future system might even be able to operate in an entirely autonomous manner. We did not dig into the engineering details necessary to build a ready-to-use system, as many issues, such as beat adjusted cross-fading, have been addressed before. Nevertheless, we believe that we could show that smartphone-based sensing has the potential to add an appealing interactive note to our future nightlife activities.

6 Acknowledgements

This work is supported under the FP7 ICT Future Enabling Technologies Programme of the European Commission under grant agreement No 231288 (SOCIONICAL). The authors would like to thank Google for providing the phones.

References

[1] R. Aylward and J. Paradiso. Sensemble: a wireless, compact, multi-user sensor system for interactive dance. In *Proc. of NIME*, 2006.
[2] L. Bao and S. Intille. Activity recognition from user-annotated acceleration data. *Pervasive Computing*, 2004.
[3] L. Bossard, M. Kuhn, and R. Wattenhofer. Visually and acoustically exploring the high-dimensional space of music. In *SocialCom*, 2009.
[4] D. Cliff. hpDJ: An automated DJ with floorshow feedback. *Consuming Music Together*, 2006.
[5] 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. *NIME*, 2008.
[6] K. Kornysheva, D. Von Cramon, T. Jacobsen, and R. Schubotz. Tuning-in to the beat: Aesthetic appreciation of musical rhythms correlates with a premotor activity boost. *Human brain mapping*, 31(1), 2010.
[7] T. Koskela, J. Julkunen, O. Kassinen, and M. Ylianttila. User Evaluation of a Community-centric Music Voting Service. *Int. Journal of Digital Content Technology and its Applications*, 4(1), 2010.
[8] J. Krakauer. Why do we like to dance – And move to the beat? *Scientific American*, 2008.
[9] M. Kuhn, R. Wattenhofer, and S. Welten. Social audio features for advanced music retrieval interfaces. In *ACM Multimedia*, 2010.
[10] K. O'Hara, M. Lipson, M. Jansen, A. Unger, H. Jeffries, and P. Macer. Jukola: democratic music choice in a public space. In *DIS*, 2004.
[11] D. Sprague, F. Wu, and M. Tory. Music selection using the PartyVote democratic jukebox. In *AVI*, 2008.
[12] T. Stiefmeier, D. Roggen, G. Ogris, P. Lukowicz, and G. Tröster. Wearable activity tracking in car manufacturing. *IEEE Pervasive Computing*, 2008.
[13] F. Styns, L. van Noorden, D. Moelants, and M. Leman. Walking on music. *Human movement science*, 26(5), 2007.
[14] L. van Noorden and D. Moelants. Resonance in the perception of musical pulse. *Journal of New Music Research*, 28(1), 1999.
[15] D. Winter. *Biomechanics and motor control of human movement*. John Wiley & Sons Inc, 2009.
[16] M. Wirz, D. Roggen, and G. Tröster. User Acceptance Study of a Mobile System for Assistance during Emergency Situations at Large-Scale Events. In *HumanCom*, 2010.
[17] M. Zentner and T. Eerola. Rhythmic engagement with music in infancy. *Proc. of the National Academy of Sciences*, 107(13), 2010.