Music Interaction on Mobile Phones

Experiment Design for Computer Science

Kuanting CHEN(201556606), Song WANG(201620694), and Meng LI(201620728)

1. Introduction

Classical music is not as popular as it was before, as suggested in a study by Kolb [4]. This study mentioned that there is a decline in audience demand for American and European orchestral music and that its audience is aging. How- ever, this study also indicated that such a declining and aging audience could be eased by increasing the probability of young people's early exposure to the music. And the study by Peery and Peery found that interacting with music can increase the understanding and appreciation of music. In the past, researchers have proposed various ways to interact with music. For example, in the system proposed by Chew et al. [3], users can adjust the tempo of the music using a driving interface. Many proposed conducting systems [1] also allow users to adjust music tempo and volume using different types of devices. Even though these systems can help users interact with music, such interactions require spe- cial devices and the systems are usually located at a fixed position. Furthermore, users must understand music very well in order to operate these conducting systems. These systems offer little help in increasing musical exposure. In order to attract most people, regardless of their abilities, the interaction approach should be the most familiar one possible and the device should be popular and nearly ubiquitous. Based on observations at many popular music concerts, we have dis- covered that the audience has a tendency to jump up and down or wave their hands to the musical beats. Several early experiments [5] have also pointed out that musical tempo can influence audience preference. This has inspired us to design an interactive interface to adjust music tempo using gestures with a smart phone.

There are numerous types and brands of smart phones, each using different devices and systems. After comparing most phone models, we determined that the acceleration sensor is essential and is embedded in almost every phone. Our proposed interface analyzes the acceleration signals to capture gestures and adjust music tempo based on the analyzed results. In order to have robust system performance, our gesture recognition method extracts the repeated intentions from the gesture signals and aligns the repeated intention with the music tempo. Our experimental results demonstrate that the proposed system can provide accurate, robust, and smooth music interaction. The user study also demonstrates that user preference to music can be improved after playing with the system. The proposed system might be helpful in promoting the popularity of classical music.

The rest of this paper is organized as follows. An overview of the system framework and detailed methods are given in Section METHODOLOGY. The empirical study is reported in Section PERFORMANCE EVALUATION. Section CONCLUSION concludes the paper and discusses future work.

2. Methodology

In our proposed system, the user can select a music piece to listen to or interact with. If the interaction mode is selected, the user should give two preparation beats for the music piece. The music then comes in on the next beat. The music tempo should follow the gesture, except in one particular scenario. If the user stops the beating gesture, the music tempo will stay the same until the end of the piece.

To achieve low user overhead, our system has two goals. First, that users can operate the system on their own without having any prior knowledge of users. Second, that users can use their own beating patterns (even for all con- ducting behaviors) when using the system. In other words, no matter if users are using simple beating gestures or very complicated conducting gestures, they can smoothly adjust the tempo. The entire system framework is illustrated in Figure 1.

As illustrated in Figure 1, our system should recognize start gestures, that is, the gesture is detected by the acceleration sensor embedded in the smartphone. The gesture data contain the x-, y-, and z-axes of the Cartesian coordinates. Initially, our system will load the input music and initialize a flag, called PlayStatus, which is set with a value equal to false if the input music has not yet been started. During the interaction mode, the system starts receiving the gesture data and waits for the start signal. Once the start gesture has been recognized, the system then sets the PlayStatus flag to be true and starts playing the input music. While the input music is playing, the gesture data will be analyzed until the end of music or until the system has been interrupted by user commands.

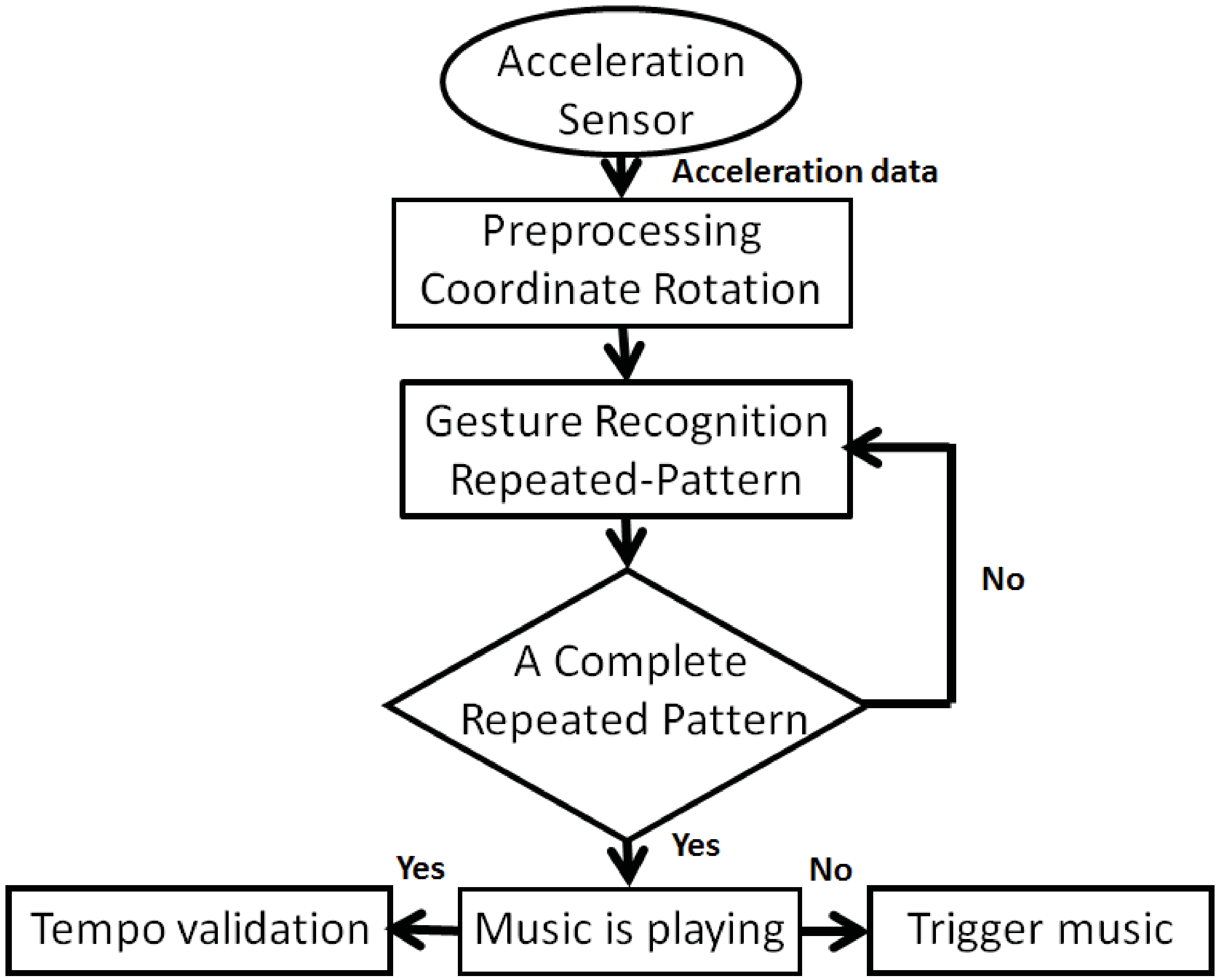

Figure 1: System framework

2.1 Data Preprocessing

Fortunately, smartphones not only contain an acceleration sensor, but they also contain a gravity and a geo-magnetic sensor, where the gravity sensor detects the direction of gravity and the geo-magnetic sensor can align the geomagnetic line. By utilizing the gravity matrix and geomagnetic sensor, the system then rotates the direction of the acceleration to fix it on the earth???s coordinates.

2.2 Gesture Recognition

Although there are various types of beating gestures, most intended beating gestures have recurring patterns to indicate beat points. Beating gestures can be categorized into numerous beat patterns, in which each fragment represents a series of hand movements through a beat. Theoretically, the lapse of time for each beat should be nearly equivalent if the tempo remains constant. Each beat pattern contains several significant features, including a start point, a beat point, and an end point. The start point and end point usually go through the origin of beating gestures, where the acceleration values of these points are zero. The beat point, which is represented by a bouncing movement, is identified by the down turning point. In other words, the beat point is the lower stationary point (e.g., the circled points in Figure 2).

If the lower stationary points are the only consideration, the inevitable hand trembles definitely affect the accuracy of beat detection. As illustrated in Fig- ure 2, the circled points are the lower stationary points. However, the green circles denote falsely detected points, which are generated from unintentional hand trembles. To avoid falsely detecting beat points, one hypothesis is made

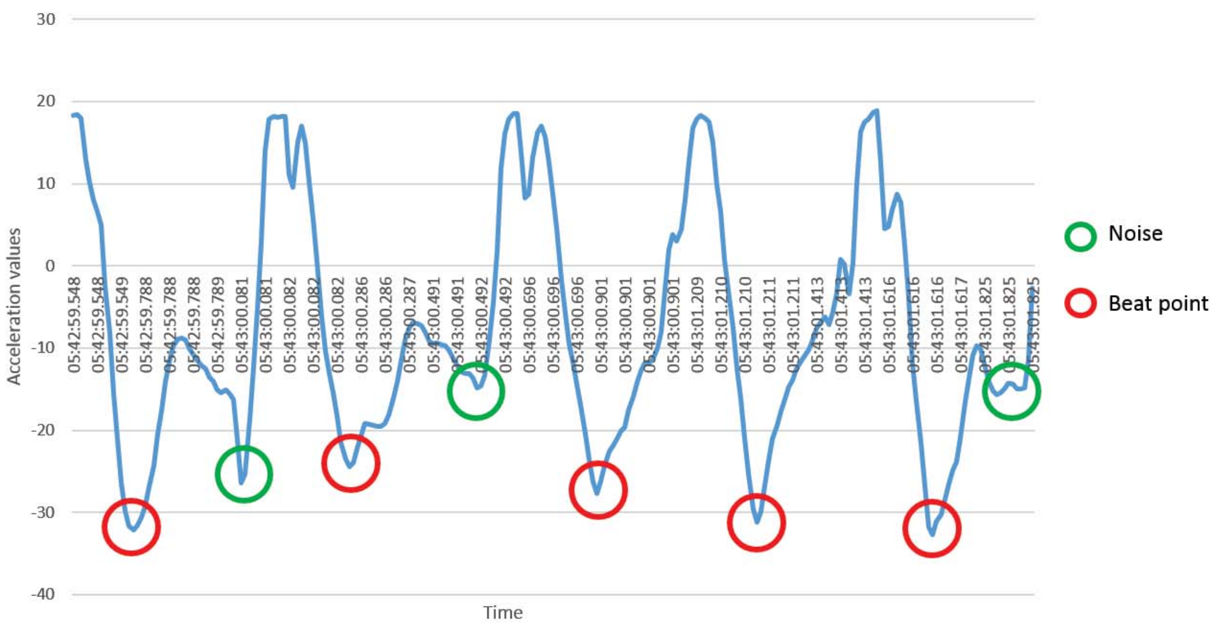

Figure 2: Samples of acceleration data

based on the study by Cambouropoulos et al. [2]. Cambouropoulos et al. stated that people prefer music with a smooth tempo. Therefore, we can assume that the tempos of two consecutive beats are nearly the same. In other words, the time periods of two consecutive beats should be nearly equivalent. Furthermore, the consecutive beating gestures should have the repeated intention. Detecting repeated intentions should be able to avoid noise and maintain tempo smoothness.

3. Performance Evaluation

3.1 Experimental Setup

There are two sets of experiments. In the first set of the experiments, data are collected from 15 users who are undergraduate or graduate students from the university. We designed a two-step test. In the pre-test, we asked the users to listen to certain classical music pieces and then to interact with any pieces they want to select. When they finished listening, we asked them about their preference regarding the music. After they interacted with music, we asked them about preference again. Half a month later, we began the post-test in which we asked the users to listen to two types of music, the same music as in the pre-test and, music with a different melody. Therefore, we can evaluate whether our music interaction system has an effect on all kinds of classical music. Finally, we asked them about their preferences regarding music pieces and the system using a 5-point scale, which indicated user preference level as very dissatisfied (1), dissatisfied, average, satisfied, and very satisfied (5). The entire experimental processes were recorded using a video camera. After interacting with the system, a thorough interview was conducted to collect user opinions.

In the first set of experiments, users select the music pieces based on their preferences. Such choices might mislead the conclusion; since users might select the songs they prefer and of course have higher preference values after interac- tion. To alleviate such concerns, the second set of experiments only allows users to interact with a selected music pieces. In the second set of the experiments, the data are collected from 25 users. Two-step tests are given. In the pre-test, users listen three classical music pieces and to interact with a selected one (i.e., Mozart K. 265/300e), Three days later, the post-test is given.

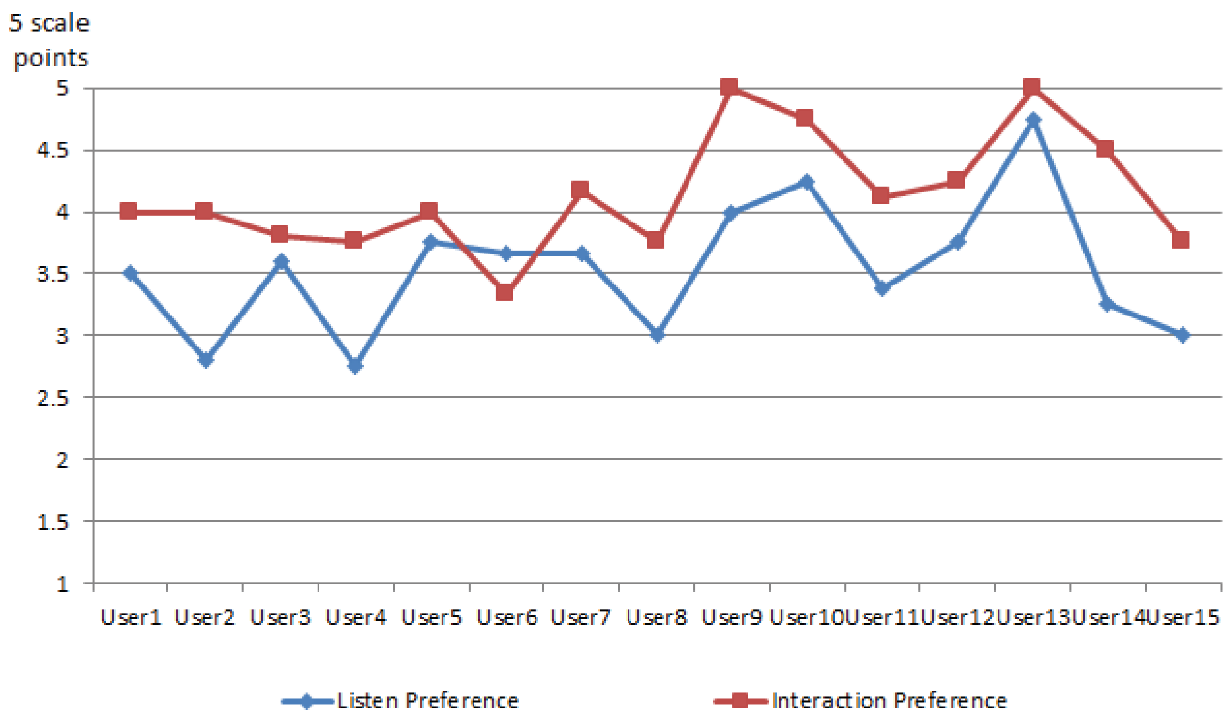

Figure 3: Users' pre-test preferences

3.2 Preference Comparison Results

Among the 15 users that participated in this set of experiments, five can play musical instruments and seven are familiar with classical music. Figure 4 illustrates that after the interaction, all subjects' preference is significantly increased except for user 6, who wanted to listen to the music only rather than using the interaction system to enjoy music. We also utilized a paired t-test (which is a statistical hypothesis test) to see whether the improvement is significant after interaction. Our hypothesis for the t-test is that preference will be increased after users interact with the system. The hypothesis passed the t-test with a p-value of $3.542 * 10^{-5}$, which indicates that the results are significant at the $p < 0.05$ level.

Even though the average preference after the post-test is higher than that of the pre-test, as illustrated in Table 4, the associated t-test shows no significant improvement after a half month. In the post-test, the p-values of the same music and the other music are 0.277 and 0.372, respectively. This may imply that period of the interaction is too short to have an impact on the users, and the interval between two tests is too long to be influenced. Therefore, the effect on preference is not statistically significant.

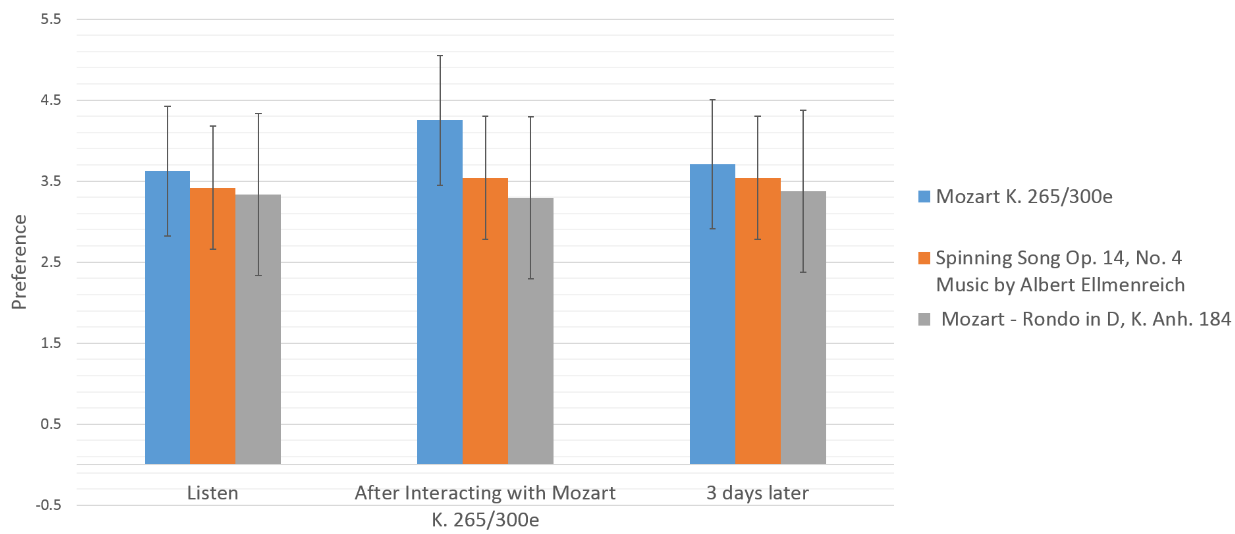

Figure 4: Preference comparison for fixed selections

Figure 5 illustrates the preference comparison for the second set of exper- iments. As illustrated in Figure 5, the average preference after interaction is significantly increased. Our hypothesis for the t-test is the preference will be increase after users interact with the system. The hypothesis passed the t-test with the p-value $5.974 * 10^{-4}$, which indicates that the results is significant at the $p < 0.05$ level. Even though the average preference of the post-test is more than that of the pre-test as illustrated in Figure 5, the associated t-test shows no significant improvement after 3 days. It may also imply that the period of the interaction is too short to impress the users and the interval between two tests is too long to remember the previous test. Therefore, the effect on the preference is not remarkable.

The improvement of preference in the post-test is not as significant as that of the pre-test, which might indicate that one can use our music interaction system more frequently to improve the long-term effect in the preference of music.

4. Conclusion

In this paper, we propose an interaction between gestures and tempo to help users enjoy music on smartphones. Given the ubiquitous nature of smartphones, most people can have the opportunity to interact with music. Our experiment shows that our system can provide an accurate, robust, and smooth musical interaction experience. In the user study, we also demonstrate that preference can be improved and that people like classical music more after using the system.

REFERENCES

1. J. Borchers, E. Lee, W. Samminger, and M. Muhlhauser. Personal orchestra: a real- time audio/video system for interactive conducting. Multimedia Syst., 9(6):594, 2004.
2. E. Cambouropoulos, S. Dixon, W. Goebl, and G. Widmer. Human preferences for tempo smoothness. In Proceedings of the VII International Symposium on Systematic and Comparative Musicology, III International Conference on Cognitive Musicology. Jyvaskyla, Finland: University of Jyvaskyla, pages 18-26, 2001.

3. E. Chew, A. Francois, J. Liu, and A. Yang. Esp: a driving interface for expression synthesis. In Proceedings of the 2005 conference on New interfaces for musical expression, pages 224-227. National University of Singapore, 2005.
4. B. M. Kolb. The decline of the subscriber base: a study of the Philharmonia Orchestra audience. International Journal of Arts Management, 2001.
5. A. LeBlanc and J. McCrary. Effect of Tempo on Children's Music Preference. Journal of Research in Music Education, 31(4):283-294, 1983.