Embodied Listening Performances Reveal Relationships Between Movements of Player and Listeners

Marc Leman Frank Desmet Frederik Styns Leon van Noorden Dirk Moelants
Dept. Of Musicology, Ghent University
Blandijnberg 2, B-9000 Ghent, Belgium
E-mail: Marc.Leman@UGent.be

Abstract

The movements of a guqin player and several listeners were recorded using kinematic sensors. Movement velocities were extracted to test whether listeners correlate with each other and with the player. The experiment revealed that listeners tend to mimic action events that underlay music. The findings provide evidence for the hypothesis that music perception has roots in action. The study offers a new methodology for studying the action-based component of music perception in several domains of music research.

1. Introduction

In embodied listening experiments, the subject is encouraged to get corporeally involved with a musical stimulus. There is evidence that music signification practices can be understood from the viewpoint of action, and that perception of music spontaneously involves action [1]. This evidence can be linked with recent theories that assume shared neuronal codes as the underlying structure for the coupling of perception and action [2, 3]. However, many questions remain unanswered, for example: how listeners signify music through action, how they engage with music, and how they engage with each other in terms of corporeal movements.

In the present study, the focus is on the measurement of corporeal movements of a single musician (called player) and a group of listeners. The aim of the paper is to explore the hypothesis in embodied music communication that movements between player and subjects, and among subjects can become synchronized. The methodology has a focus on movement patterns and comparison of movement patterns among different subjects and different listening sessions.

In the present paper, we use movement infra red camera capturing systems and joy-stick sensor technology to monitor (a) how a musician moves while playing and (b) how listeners move while attuning or harmonizing with the perceived music. From the measurement of these movements, we extract information about the origin and the nature of the emulated behavior using standard multivariate statistical techniques. Velocity is a core cue of movement, for that reason we focused on the extraction of movement velocity. This focus is in line with several other studies in which movement velocity has been used as a cue.

Figure 1. Player and motion capturing

2. Experimental Setup

In order to identify the movements that underlay the actions of player and listener, audio, video and kinematic data of three short (<30 sec) pieces (P1,P2,P3) of Chinese guqin music were recorded. Two of these pieces, P1 and P2, have a rather fluent melodic line which is clearly structured. In contrast, P3 has a more rubato and narrative character with a less fluent melodic line. The kinematics of the player was recorded with an infra red camera system (Figure 1). Likewise, thirty subjects, unfamiliar with this music, and with Chinese music in general, were asked to

Proceedings of ENACTIVE/07
4th International Conference on Enactive Interfaces
Grenoble, France, November 19th-22nd, 2007

move a joystick while listening to paired repetitions of the audio recordings (Figure 2).

Figure 2. Listener responding to music via joystick

This task was done in four sessions that immediately followed each other and induced learning. In the first session, subjects were asked to move the joystick while listening to the musical audio stimulus. Before the start of the second session, subjects were invited to listen to the fragments in random order and as many times as they wanted. When finished, they were asked to move to the same sequence of musical fragments as in part one. Before the start of the third session, subjects were asked to listen to the musical fragments, now using a visual representation (pitch versus time graphs) of each musical fragment. Again the subjects were allowed to listen/look in random order as many times as they wanted. When finished, they were asked to move as in session one. Before the start of the fourth session, subjects were asked to watch/listen a video of the player's performance of the three pieces. The subjects could watch/listen in random order and as many times as they wanted. When finished, they were asked to move as in session one.

The listener's kinematics were recorded and after each session, the listeners had to fill out a questionnaire that probed the subject's self-assessment of the performance. The main question was “Do you have the feeling that your movements were in accordance with the music?” (‘very badly’ to ‘very well’, asked in session 1 to 4, for each piece). All subjects were thoroughly verbally instructed about what they were expected to do at the beginning of the experiment. The subjects were also given a short written instruction that repeated what the instructor said.

3. Data Analysis

Movement velocity is calculated as the displacement, in three dimensions (player) or in two dimensions (listener), during non-overlapping time windows of 250 ms. Movement velocities are used in two types of measurement.

First, intra-subject comparisons address the subject's ability to repeat the movement velocities during subsequent repeated listening of a piece (e.g. P1P1) during a session. A correlation between both movement velocity patterns at the $p < 0.01$ significance level was used as a criterion for making the distinction between successfully and non-successfully repeated movements. This choice was based on a detailed inspection of a number of correlated movement velocities.

Second, inter-subject comparisons access the relationship between the movement velocities of different listeners. Similarly, the computational comparison of the movement velocity vectors is based on correlation. In addition, for inter-subject comparisons, a dynamic time warping technique is used. In the latter, the listener's movement velocities are first warped onto a reference pattern, namely, the player's movement velocities. Correlation analysis is then based on warped patterns. This dynamic time warping (DTW) accounts for the fact that movements of the subjects can be somewhat shifted as compared to the player 's movements (anticipated or lagged).

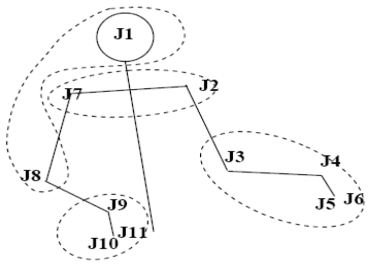
Moving along with music may be *in tune* or *out of tune*. In tune would mean that the listener has the feeling that the movements somehow are in accord with the music. Out of tune would mean that they are not in accord. The latter would occur, for example, when no movements were made, when hesitating movements were made, or when the listener had the feeling that the movements were too late or too early, and so on. In order to be able to address the action component in relation to music, it was important to try to access this distinction in an objective way. To address this issue in a quantitative way, it was assumed that listeners who found a proper motor strategy to express their movement *in tune* with music (during 30 seconds) would not radically change their strategy in a second performance that immediately followed the first one, while listeners, who found that their motor articulations were *out of tune* with music would be inclined to change strategy and therefore have a different performance in the second trial. Thus, high correlation between successive performances (e.g. P1P1) would indicate *reliable performance*, while low correlation would indicate *non-reliable performance*. The measurement of the correlation between the two repetitive performances thus offers an objective criterion for analysis.

4. Results

In the course of this study, there were no effects of gender, age, musical background, listening hours per week, and absolute pitch perception. Therefore, these effects can be discarded from further analysis.

Dynamic time warping (DTW) was used to compare movement velocities of the listeners with movement velocities of all joints of the player. All reliable performances were selected for this DTW analysis. The warp path was restricted to a maximum of 750 ms from the diagonal (which corresponds with 3 time units). The path with the best resulting correlation and the lowest cost was chosen. The resulting cost/correlation analysis shows that there is a relation between correlation and cost. The higher the DTW correlation obtained, the lower the cost it takes to achieve this. This analysis shows that the movement of the right shoulder of the player has the highest average correlation with the lowest cost (Figure 3b). There is also an indication that the player's movements can be classified into 4 different groups: right hand; left elbow and left hand; head and right elbow; shoulders (Figure 3a).

(a)



(b)

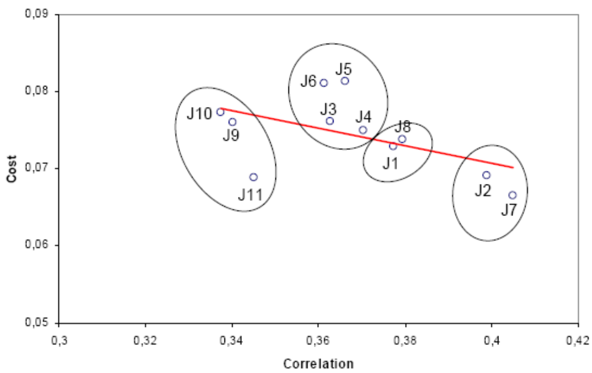


Figure 3. (a) Grouping of joints of the player, (b) Correlation and cost between different joints of the paper and the response of the listeners

Based on the above results, the movement velocities of the listeners were then time-warped with one single movement marker, the right shoulder. General Linear Modelling (GLM) for Repeated Measures was then used to model the number of significant listener-player correlations for *reliable* and *non-reliable* performances over the four sessions. For the reliable performances, there is a significant effect over session ($p = 0.005$) and no significant interaction effect between session and piece. The latter means that the correlation of the movements with those of the player increase over session and that the pieces follow different trends. In contrast, for the non-reliable performances there is no significant effect over session. This result for non-reliable movements confirms the separation between the two types of movements (reliable versus non-reliable).

When listeners move in tune with the music, one may expect that inter-subject correlations can be observed. In the present study, the number of inter-subject correlations, based on all performances (reliable as well as non-reliable) was low. GLM for Repeated Measures reveals a significant difference between pieces. This difference can be attributed to the difference between P1, P2 on the one hand and P3 on the other hand. The analysis shows a significant effect over session ($p < 0.001$) and a significant interaction between session and piece ($p < 0.001$). Data analysis based on warped movements shows inter-subject correlations in about 50% of the movements for P1 and P2, and about 30% for P3. The rubato narrative character of P3 seems to be less predictable in terms of action events, and the movements seem to have a less pronounced inter-subjective component.

Concerning intra-subject correlations, the number of reliable performances increases over sessions. GLM for Repeated Measures reveals a significant difference between pieces ($p < 0.05$). There is a significant effect of session for a linear model ($p = 0.005$) and there is no significant interaction between session and piece. Learning works best for P2 and P1, that is, from above 40% in S1 to about 70% in S4. Instead, for P3, there is a smaller increase. This means that, for reliable performances, subjects tend to improve in their individual learning path for all pieces. Seen in the context of the inter subject results, it can be concluded that this learning effect is based on individual interpretations rather than on common inter-subjective interpretations. In other words, over sessions, the individual interpretations improve, but they drive away from a common inter-subjective interpretation

5. Discussion

A direct comparison of the movement velocities of the player with the movement velocities of listeners revealed some interesting relationships.

First, it was observed that the movement velocities of the listeners' arm movements tend to correlate with the movement velocity of the player's shoulders. This finding seem to imply (i) that listeners can embody the musical stimulus by translating the perceived sounds into actions, and (ii) that action components of listeners are related to action components of the performer. It should be noted here that guqin music is particularly suited for studying these relationships because the technological mediation between the player's bio-mechanical energy and the sound energy is very straightforward. Due to the fact that the instrument has neither frets nor bow involved, the sonic trace in music can be considered a direct image of the player's action. Therefore, it could be argued that the sonic traces of the guqin facilitate the transformation into action patterns that relates to the player's action pattern. However, listeners do not emulate pitch modulation movements because if they did, there would be a correlation with the movements of the player's finger. Instead, it was observed that listeners tend to emulate the player's shoulders. The role of the shoulders is particularly striking and it may point to the fact that listeners emulate the player's musical *intention* rather than the player's technical playing gestures. The decoding can only be explained by the fact that listeners have a capacity to emulate the action patterns that are contained in music. The findings suggest that through action, the listener builds up a corporeal understanding of the musical patterns. Music, under certain circumstances, would thereby operate as a channel through which action patterns can be transmitted.

Second, it was observed that the music-driven movement velocities of the listeners tend to correlate with each other. Beat and rhythm are often concerned as the main sources for musical embodiment; however, the musical stimuli used in the present study did not contain a very clear beat. Despite these conditions, a growth of correlation among the movement velocities of the listeners could be observed over sessions. Therefore, it can be concluded that listener-listener correlations may be based on the decoding of expressive features in music, rather than on mere beat following. The present study suggests that embodied listening is prone to learning and that in a condition where learning is stimulated, the quality of the listener's movements tend to improve. This learning, apparently, may be strongly depending on the musical character, as the difference between P1-P2 and P3 seem to indicate.

Finally, it was noticed that Piece 3 revealed an interesting difference between intra and inter subject analysis. The intra subject analysis suggests that listeners tend to improve for P3 over session, while the inter subject analysis suggests that listeners tend to drive away from a common interpretation. The two trends can be explained by the fact that listeners develop their own embodied listening solution to P3 in a consistent way but this listening solution differs among listeners. The lack of a shared inter subjective embodied listening solution may be due to the narrative character of the piece.

To sum up, the present study suggests that embodied listening is based on a music communication model which consists of (i) the sonic encoding of actions by the player, (ii) the transmission of these sonic cues (through music), and (iii) the decoding (or corporeal understanding) in terms of action events by listeners who perceive these sonic cues. The transmission of a corporeal code forms part of a mechanism where listeners, in order to bridge the semantic gap between sonic cues and meaning, attribute ‘plausible’ or ‘virtual’ action events to music, which they take from their own action repertoire. In this way, through embodied attuning, music perception is grounded in the listener's subjective experience. This foundation in action will give it a proper inter-subjective aspect, as shown in this study.

6. Conclusions

The present study provides evidence for the theory that music perception is rooted in action. Based on direct measurement of movements, it was shown that listeners tend to mimic action events that underlay music. Embodied listening can be explained by the existence of shared codes for action and perception, allowing the translation from sonic energy via bio-mechanical energy into the perception of action, or action intentions. Clearly, the decoding of action in music draws on a subtle mechanism in which the listener tends to disambiguate the stimulus by body movement. The present study suggests that this mechanism is not necessarily based on knowledge of how to play the instrument, nor is it merely constrained by the bio-mechanics of the human body. It was observed that the (reliable) movements of the listeners' arms correlate with the movements of the player's shoulder. This seems to indicate that there can be a genuine transfer of action patterns from the player to the listeners. “If this hypothesis is true, then it points to the existence of a corporeal code which can be encoded and decoded and which music is able to transmit.

References

[1] Leman, M., Embodied music cognition and mediation technology, Cambridge, The MIT-Press, in press

[2] Hommel, B., Musseler, J., Aschersleben, G., and Prinz, W., “The Theory of Event Coding (TEC): A framework for perception and action planning”, Behavioral and Brain Sciences, 24(5), pp. 849-939, 2001.

[3] Sebanz, N., Bekkering, H., and Knoblich, G., “Joint action: bodies and minds moving together”, Trends in Cognitive Sciences, 10(2), pp. 70-76, 2006.