

SHARING MUSICAL EXPRESSION THROUGH EMBODIED LISTENING: A CASE STUDY BASED ON CHINESE GUQIN MUSIC

MARC LEMAN, FRANK DESMET, FREDERIK STYNS,
LEON VAN NOORDEN, AND DIRK MOELANTS
Ghent University, Ghent, Belgium

IN THIS STUDY WE REPORT ON THE RESULT OF AN experiment in which a guqin music performance was recorded and individual listeners were asked to move their arm along with the music that they heard. Movement velocity patterns were extracted from both the musician and the listeners. The analysis reveals that the listeners' movement velocity patterns tend to correlate with each other, and with the movement velocity patterns of the player's shoulders. The findings support the hypothesis that listeners and player share, to a certain degree, a sensitivity for musical expression and its associated corporeal intentionality.

Received May 8, 2008, accepted November 3, 2008.

Key words: embodied music cognition, music perception, expression, guqin music

MUSICAL SIGNIFICATION PRACTICES INVOLVE a high degree of corporeality and, therefore, it is interesting to study music perception from the viewpoint of body movement. In a recent survey by Lesaffre et al. (2008), 95% of participants ($n = 663$) reported that they move spontaneously along with music while listening. Apparently, when people enjoy music, there is often body movement that goes along with it.

In the past, music perception often has been linked with movement. For example, Aristotle assumed that music is reflecting "men in action," and that music, through melody and rhythm, expresses different characters with which our soul may attune during the perception of music (Aristotle & McKeon, 2001). Hume suggested that a kind of sympathy is essential in our perception of art (Hume, Selby-Bigge, & Nidditch, 1975), while Kant (1790/1924) introduced the notion of *sensus communis* as a kind of intersubjective sensitivity for beauty in art. In the late 19th century,

Hanslick (1891) introduced the idea that the perception of moving forms ("tönend bewegte Forme") is sufficient for signification. However, it was Lipps (1903) who, in a more explicit way, spoke about the corporeal articulation of expression ("Ausdrucksbewegungen"). Through moving along with the forms in art (with our body and/or our mental imagery), he claimed it is possible to enter into an empathic relationship with (musical) forms and, consequently, to fully signify them. Later on, steps were undertaken to identify different types of body movement in relation to music perception, such as expressive movements (Truslit, 1938; see also Repp, 1993), repetitive movements related to conducting (Becking, 1928; see also Nettheim, 1996), or movements that express emotional content (Clynes, 1977).

In recent work, the relationship between music perception and movement has been studied from different viewpoints, such as musicology and philosophy (e.g., Cummings, 2001; Hatten, 2003), behavioral studies about music performance, walking on music, activation of emotions (e.g., Friberg & Sundberg, 1999; Friberg, Sundberg, & Fryden, 2000; Gabrielsson, 2003; Juslin & Laukka, 2004; Palmer 2005; Styms, Van Noorden, Moelants, & Leman, 2007; Sundberg, 2003), cognitive neuroscience (e.g., Brown, Martinez, & Parsons, 2006; Chen, Zatorre, & Penhune, 2006; D'Ausilio, Altenmüller, Belardinelli, & Lotze, 2006; Zatorre, Chen, & Penhune, 2007), and computational modeling (e.g., De Poli, 2004; Camurri, Volpe, De Poli, & Leman, 2005; Mion & De Poli, 2008; Widmer & Goebel, 2004). Movement is now seen as an important factor that contributes to perceptual disambiguation (Phillips-Silver & Trainor, 2008). Movement simulation also accounts for the better recognition of and better synchronization with one's own audio-recorded piano performance (Keller, Knoblich, & Repp, 2006; Repp & Knoblich, 2004). The link between music perception and movement has furthermore been supported by evidence for shared neuronal codes of perception and action (e.g., Berthoz, 2000; Hommel, Musseler, Aschersleben, & Prinz, 2001; Jeannerod, 1994; Wilson & Knoblich, 2005), with important implications for our understanding of

social musical interaction (Clayton, 2007; De Bruyn, Leman, Demey, & Moelants, in press; De Bruyn, Leman, & Moelants, 2008; Keller, 2008).

In short, in our understanding of music perception, movement often has been at the focus of attention, through notions that imply the sharing of musical expression and the coupling of perception and action. Thanks to recent developments in technology a direct study of expressive body movement in relation to music perception now is possible (Leman & Camurri, 2005), although measurement and direct evidence still is hindered by a number of methodological difficulties, such as the identification of reliable movements, and their distinction from unreliable or arbitrary movements (Wanderley, Vines, Middleton, McKay, & Hatch, 2005). In addition, there are particularities of data analysis that deserve special attention, because movement data are time series and the data-analysis of time series is known to be difficult especially when human movement is involved (Desmet, Leman, Lesaffre, & De Bruyn, in press).

The aim of the present paper is to explore whether music expression can be studied and understood within the paradigm of embodied music cognition (Leman, 2007), using empirical methods that comply with the particularities of music-driven human movement. In the embodied music cognition approach it is assumed that the human body is the natural mediator between experience and physical reality. Music-driven movements of the listener are seen as corporeal articulations that reveal a mirroring (in movement) of the perceived sonic moving forms. Through this mirroring, it is assumed that musical expression can be captured more fully, experienced, and understood.

Correlations of music-driven movements among listeners would support the hypothesis that their embodied perception of music expression is shared. Correlations between listeners and the player would further support the hypothesis that there is a link between what listeners perceive as expression, and what the player produces as expression. A correlation would support the idea that listeners can somehow decode (and reveal through music-driven corporeal articulations) the expression that the player has encoded in the musical audio.

A major research question is how the above hypothesis of embodied music cognition can be tested and further explored. In the present paper, we use fragments of Chinese guqin music as stimuli in an experiment that measures and compares music-driven corporeal articulations.

The guqin (pronounced *ku-chin* in English) is interesting for our purpose. The instrument belongs to the

family of the zither, a plucked stringed instrument. It consists of a long and narrow hollow wooden box that functions as a sound box and whose upper part functions as a fretless fingerboard on top of which there are seven strings attached, each about 110 cm long. The strings are tuned to a pentatonic scale. The way in which guqin music is played, namely by plucking the string with the right hand and moving the finger of the left hand over the string, makes it suitable for a detailed study of the relationship between movements of listeners and movements of the player. Indeed, the guqin has no bow, and fingering is directly related to sound, as there are no frets to interfere within the sliding. Although the playing technique is rather complex, the sliding tones in guqin music can be conceived as sound patterns that reflect aspects of the playing movement, without much intermediate technology. Guqin music is thus of particular interest to embodied music cognition research because the encoding of playing gestures into sound patterns proceeds in a relatively direct way, which facilitates the study of possible relationships between movement and sound (see Li & Leman, 2007).

The methodology that we developed for this study has a focus on the registration and extraction of the movement velocity. The data analysis is based on a technique that combines the correlation of time series with cost functions that account for local time shifts of the recorded time series (Desmet et al., in press).

Method

In this study, we made an audio, video, and movement recording of three fragments of guqin music. Then we asked listeners to listen to these fragments and respond by spontaneously moving a telescopic stick along with the music. The movements of the listeners were recorded.

Stimuli

The musical stimuli used in this experiment were three musical fragments (P1, P2, P3) of the traditional Chinese piece “Missing an old friend” played on the guqin by Henbing Li, an experienced guqin player educated at the Central Conservatory of Music in Beijing. Each fragment had a duration of about 30 s. Of the three musical fragments, P1 and P2 have a rather fluent melodic line that is clearly structured. In contrast, P3 has a more narrative character with a less fluent melodic line. The fragments contain pitch slidings (in glissando and vibrato patterns), which is typical for guqin music. None of the fragments has a clear regular beat. The audio and video of this performance was recorded on

hard disk.¹ In addition, a motion capturing system with infrared cameras was used to monitor the movements of 11 different markers attached to the joints of the musician (Figure 1, see color plate section). The movements were recorded at a sampling frequency of 100 samples/s using the Qualisys Motion Capture System (www.qualisys.com).

Task

During the experiment, listeners were asked to move a joystick while listening to the music through headphones. Each fragment was preceded by five short beeps in order to announce the next fragment. The joystick was attached horizontally to a stand. A telescopic stick, which allowed free movement of the arm in a considerable space in front of the participant, was attached to the joystick and participants were asked to move this stick with the right or left arm (see Figure 2 in color plate section). The adapted joystick (Saitek ST 90) was connected to a computer via USB, using the musical software AnalogBox (www.andyware.com). The movements were recorded in a two-dimensional plane, using horizontal and vertical coordinates. The movement data were stored on hard disk at a sampling frequency of 100 samples/s. After each session, listeners were asked to fill in a questionnaire where they gave an assessment of their performance.

Participants

Thirty participants (13 men, 17 women) participated in the experiment. They were selected from a database of participants that was set up in the context of a musical audio mining project (MAMI) at Ghent University (Lesaffre et al., 2008). None of the participants knew Chinese guqin music. All the participants explicitly agreed that their data might be used for scientific research. The age of the participants varied from 21 to 35 years with an average of 27.6 years. Fifteen participants played an instrument. On average, the participants listened to music 19.6 hours ($SD = 13.1$) per week. Participants were paid 10 Euro for their participation to the experiment.

Procedure

Participants were asked to fill in a questionnaire about gender, age, musical background, listening hours per week, absolute pitch perception. In order to become familiar

with the equipment before the start of the experiment, participants were asked to move the joystick while listening to a musical fragment by Chopin (Preludes in C major—24 Preludes Opus 28). The fragment was chosen to acquaint participants with the joystick and the task of moving along with music. Participants could practice as many times as they wanted. The experiment itself then consisted of four sessions that followed each other. In between sessions, participants received information about guqin music. After the first session, participants could listen to the music as much as they wanted. After the second session, participants could inspect a graphical score, and after the third session, participants could watch the video of the player. After each session, participants reported on their own performance (the processing of these data is not contained in the present paper). The four sessions enter into the analysis in order to check consistency and possible learning effects.

Each session contained the same three pieces (P1, P2, P3) of guqin music, played in the same order. The participant had to listen and move to the same piece twice in a subsequent repeated trial as in: P1-P1—P2-P2—P3-P3. The repeated piece was always separated by a short break. Between pieces, a slightly longer break was used. Each participant took part in 12 trials (three pieces in four subsequent sessions).

Data Analysis

The movement velocity pattern is a time series whose values give the amount of displacement over subsequent time units. Given a lack of frequencies above 2 Hz, the sampling rate of the movement velocity patterns could be reduced to 4 samples/s, which corresponds to subsequent time units of 250 ms. The displacement is based on three dimensions in case of the infrared camera system, and on two dimensions in case of the joystick. However, in the movement velocity pattern, three or two dimensions are reduced to one single dimension, so that a comparison between movements recorded with both systems becomes possible.

In calculating the correlation between any two movement velocity patterns, both the raw and the warped patterns are studied. Raw patterns are just the original movement velocity patterns, while warped patterns are based on dynamic time warping (see Appendix), which is a technique that allows local time shifts of values in time series. The warped time series compensates for the fact that movements of the listener can be locally time shifted as compared to the player's movements, due to anticipation or delay. In this study, the dynamic time warping envelope was restricted to a maximum of 750 ms

¹See www.ipem.ugent.be/2005/Guqin/video/P1P2P3.wmv

(which corresponds with three movement velocity samples). Figure 3 (see color plate section) gives an example of the dynamic time warping technique. Correlation of warped movement velocity patterns can produce better results than correlation of raw movement velocity patterns because there can be a better match between the peaks of the patterns. However, this improvement is not guaranteed (see Appendix). The number and nature of the local shifts can be calculated as a cost of the dynamic time warping, which can be taken into account in the analysis. For example, in the section on listener-player correlations, this cost is taken into account in order to determine the body part of the player that best correlations with the arm movements of the listener.

Data analysis is based on intralister correlations, interlister correlations, and listener-player correlations: (i) Intralister correlations address each participant's ability to repeat movement velocities during a subsequent repeated listening of a piece (e.g., as in P1-P1 or P2-P2). The number of significant correlations ($p < .01$) reflects the extent to which the population can perform this task in a reliable (that is, repeated) way. For that reason, these performances are also called *reliable* performances. The choice for $p < .01$ as a criterion for making the distinction between reliable performances and unreliable performances was based on a detailed inspection of the number of correlated movement velocities. (ii) Interlister correlations address the relationships among the movement velocity patterns of listeners. (iii) Listener-player correlations address the correlations between movement velocity patterns of listeners (one variable) and the movement

velocity patterns of the different joints of the player (11 variables).

Results

A repeated measures ANOVA revealed that there were no effects of gender, age, musical background, listening hours per week, and absolute pitch perception. This is in agreement with the fact that all participants were unfamiliar with the music. Therefore, these effects were discarded from further analysis.

Intralister Correlations

ANALYSIS OF RAW MOVEMENT PATTERNS

Figure 4 shows the relative amount of reliable performances as a mean value for all participants, using the unwarped patterns. The amount increases for the three fragments P1, P2, and P3 over the sessions S1, S2, S3, S4. This increase works best for fragment P1 and fragment P2, where it goes from above 40% in session S1 to about 70% in session S4. Instead, for fragment P3, there is only a slight increase, as the number goes from 40% to about 45%. The optimal fit was calculated by selecting the maximum R^2 value for different regression types. The best fit was obtained using a second order regression. GLM (General Linear Modeling) for Repeated Measures (Tables 1a and 1b) revealed a significant difference between pieces ($p < .05$), a significant effect of session for a linear model ($p = .005$) and no significant interaction between session and piece. This means that, for reliable performances, participants tend to improve in their performance for all pieces over sessions.

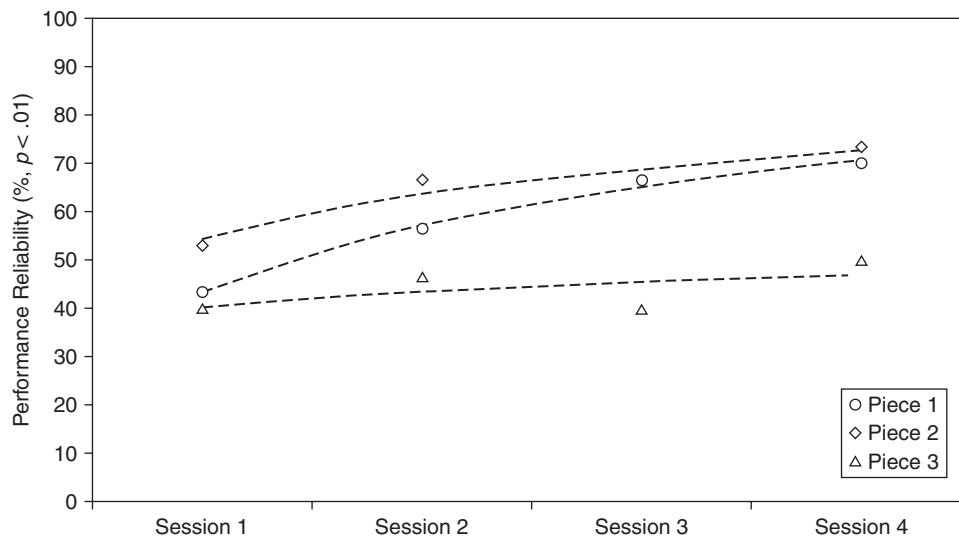


FIGURE 4. The relative amount of reliable performances as a mean value for all participants, using the unwarped patterns.

TABLE 1. GLM for Repeated Measures Applied to the Data of Figure 4.

(A) Test for Differences Among Pieces							
Source	Type IV Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	<i>p</i>	Partial Eta Squared	
Intercept	113.34	1	113.34	278.69	.000	.76	
Piece	2.77	2	1.39	3.41	.04	.07	
Error	35.38	87	0.41				

(B) Test for Differences Among Sessions							
Source	Session	Type IV Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	<i>p</i>	Partial Eta Squared
Session	Linear	1.50	1	1.50	8.14	.005	.09
	Quadratic	0.04	1	0.04	0.25	.62	.003
	Cubic	0.11	1	0.11	0.56	.46	.006
Session* Piece	Linear	0.33	2	0.17	0.91	.41	.020
	Quadratic	0.07	2	0.04	0.20	.82	.005
	Cubic	0.09	2	0.04	0.23	.80	.005
Error (session)	Linear	16.06	87	0.19			
	Quadratic	15.38	87	0.18			
	Cubic	16.90	87	0.19			

ANALYSIS OF WARPED MOVEMENT PATTERNS

The results obtained with the dynamic time warping technique applied are shown in Figure 5, and Table 2a and 2b (for the GLM analysis). The number of reliable performances over session is from 65% to 85% for P1, and from 60% to 70% for P2. With warped data, P3 gets 35% to above 40%. The results with warped data are not very different from the results with unwarped data,

suggesting that participants do not anticipate or delay their own performances. In both cases, the major result is the number of reliable patterns, which indicates that listeners can repeat their own movements when listening to the same piece that is immediately repeated. However, this number depends on the character of the piece, as the differences between P1 and P2 with respect to P3 clearly illustrate.

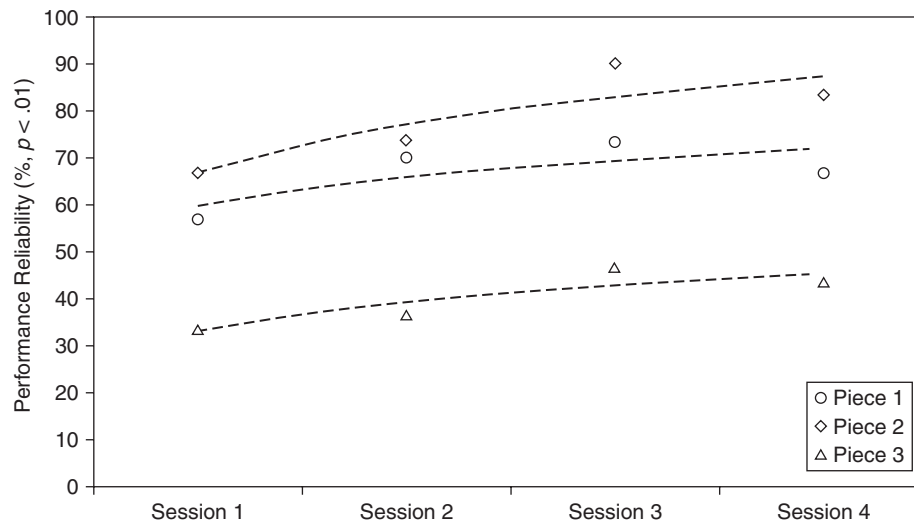


FIGURE 5. The relative amount of reliable performances as a mean value for all participants, using the warped patterns.

TABLE 2. GLM for Repeated Measures Applied to the Data of Figure 5

(A) Test for Differences Among Pieces							
Source	Type IV Sum of Squares	df	Mean Square	F	p	Partial Eta Squared	
Intercept	138.14	1	138.14	441.16	.000	.84	
Piece	9.37	2	4.69	14.97	.000	.26	
Error	27.24	87	0.31				
(B) Test for Differences Among Sessions							
Source	Session	Type IV Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
Session	Linear	0.93	1	0.93	15.38	.002	.06
	Quadratic	0.47	1	0.47	12.69	.01	.03
	Cubic	0.09	1	0.09	1.51	.05	.006
Session* Piece	Linear	0.11	2	0.05	0.31	.73	.007
	Quadratic	0.11	2	0.05	0.30	.74	.007
	Cubic	0.15	2	0.07	0.40	.67	.009
Error (session)	Linear	15.11	87	0.17			
	Quadratic	15.18	87	0.17			
	Cubic	16.11	87	0.19			

Interlistener Correlations

The number of interlistener correlations based on the raw data of all reliable ($p < .01$) performances were low (Figure 6), but with warped data interlistener correlations were obtained in about 50% of the movements for fragments P1 and P2, and about 30% for fragment P3 (Figure 7). This seems to indicate that

movement velocities of different participants display some anticipation and delay. GLM for Repeated Measures (Table 3a, related to Figure 7) revealed a significant difference between pieces ($p < .01$). In addition, the analysis showed a significant effect over session ($p = .002$) and a somewhat weak interaction between session and piece ($p < .01$) for a cubic model. The latter may be attributed to the slightly deviant

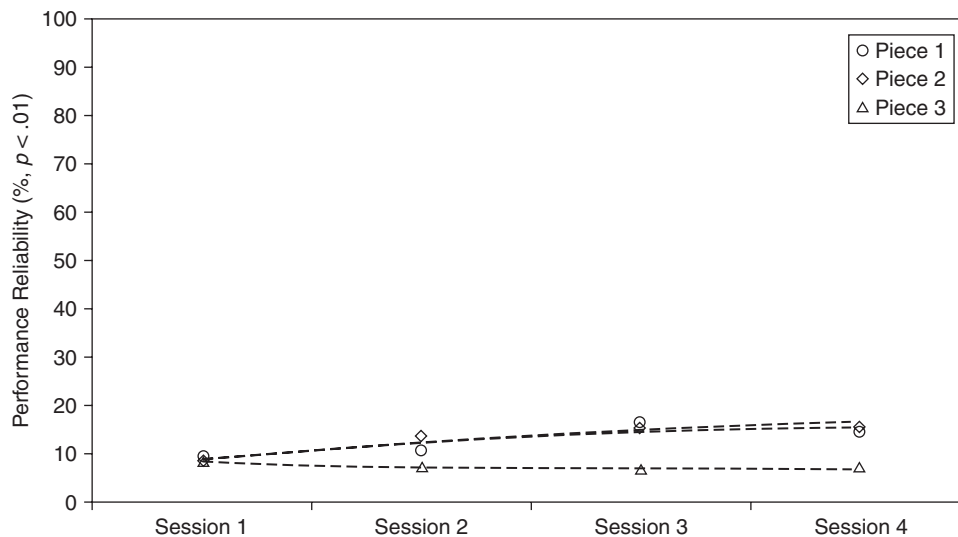


FIGURE 6. The number of interlistener correlations, based on all reliable ($p < .01$) performances, unwarped.

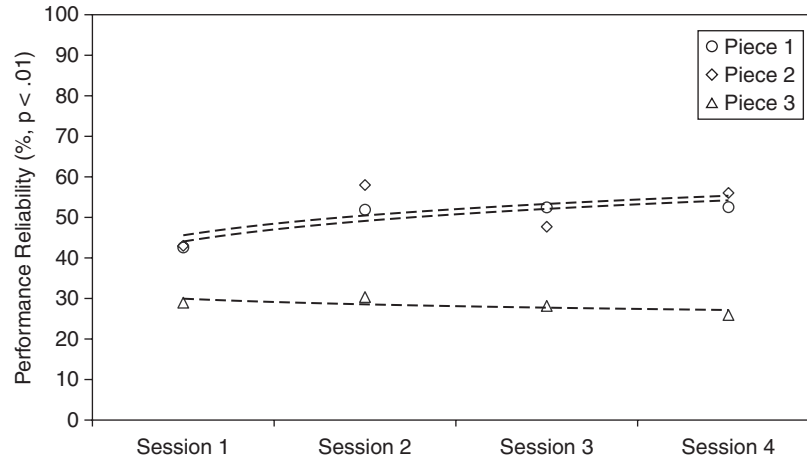


FIGURE 7. The number of interlistener correlations, based on all reliable ($p < .01$) performances, warped.

trend of fragment P3, compared to fragments P1 and P2. The rubato narrative character of fragment P3 seemed to have a less pronounced degree of interlistener agreement, but a higher degree of personal interpretation as sessions proceeded.

Listener-Player Correlations

In studying the correlation between listener and player we used the movement data from each listener and

compared these with the movement data of the different joints of the player. The joints of the player were labeled as follows: J1 = head, J2 = left shoulder, J3 = left elbow, J4 = left wrist, J5 = left thumb, J6 = left ring finger, J7 = right shoulder, J8 = right elbow, J9 = right wrist, J10 = right middle finger, J11 = right index finger (see Figure 8). In this comparison, we used the dynamic time warping technique by considering the total number of local shifts as a cost. This cost was then taken into account with the correlation analysis.

TABLE 3. GLM for Repeated Measures Applied to the Data of Figure 7.

(A) Test for Differences Among Pieces

Source	Type IV Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
Intercept	225300.10	1	225300.10	172.73	.000	.67
Piece	13270.02	2	6635.01	5.09	.008	.11
Error	113476.38	87	1304.33			

(B) Test for Differences Among Sessions

Source	Session	Type IV Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
Session	Linear	386.42	1	386.42	3.88	.05	.04
	Quadratic	321.11	1	321.11	3.32	.07	.04
	Cubic	520.57	1	520.57	10.61	.002	.11
Session* Piece	Linear	564.67	2	282.34	2.84	.06	.06
	Quadratic	40.11	2	20.05	0.21	.81	.005
	Cubic	492.54	2	246.27	5.02	.009	.10
Error (session)	Linear	8659.21	87	99.53			
	Quadratic	8420.28	87	96.79			
	Cubic	4268.59	87	49.06			

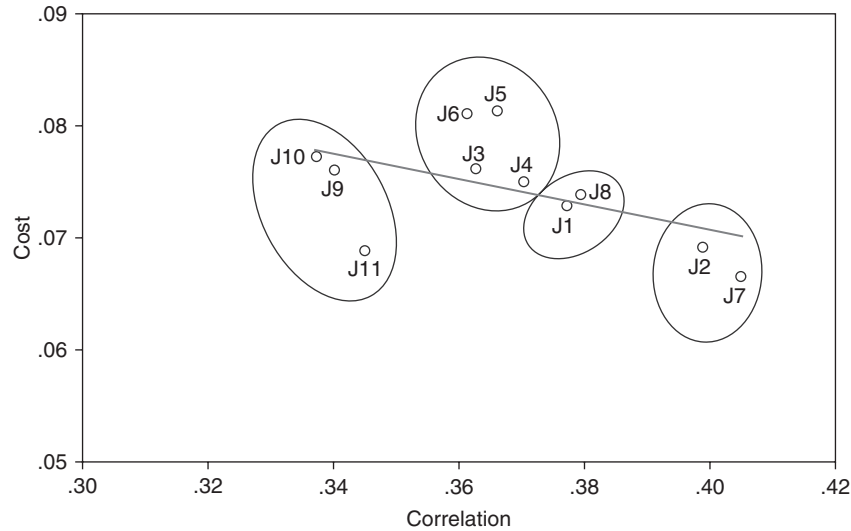


FIGURE 9. Correlation and cost between (i) different joints of the player (see text for their meaning) and (ii) the arm response of the listeners.

Figure 9 shows that there is an interesting relation between correlation and cost. The higher the correlation obtained (horizontal axis), the lower the cost it takes to achieve this (vertical axis). This analysis therefore clearly shows that the movement velocity of the right shoulder of the player (J7) has the highest average correlation and the lowest cost compared with the arm movement of the listeners. Based on the movement velocity patterns of 11 different joints, the player's movements can be classified, using cladogram analysis, into four different groups, namely right hand, left elbow and left hand, head and right

elbow, and shoulders (Figure 9). The cladogram analysis (using Phylip, <http://evolution.genetics.washington.edu/phylip.html>) aimed at organizing a given set of movement data according to character state changes in the corresponding encoded sequences. Based on the above results, we selected the right shoulder (J7) as the reference movement that will be used for further comparison with movements of the listeners.

The average number of correlations between unreliable performances and the player's right shoulder (J7) are below 40% for all fragments (Figure 10). GLM for

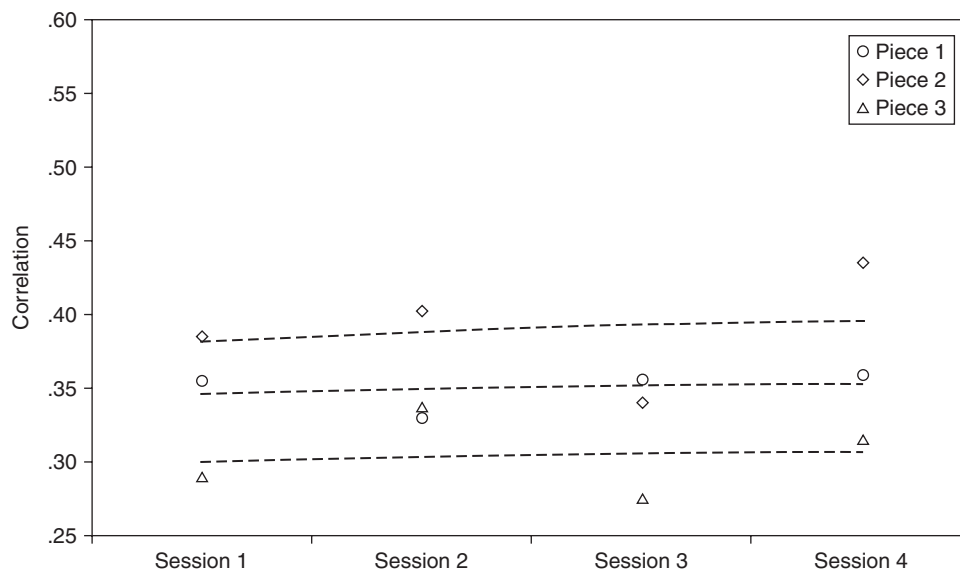


FIGURE 10. Average of correlations between listeners' movements and J7 for *unreliable* performances over the four sessions.

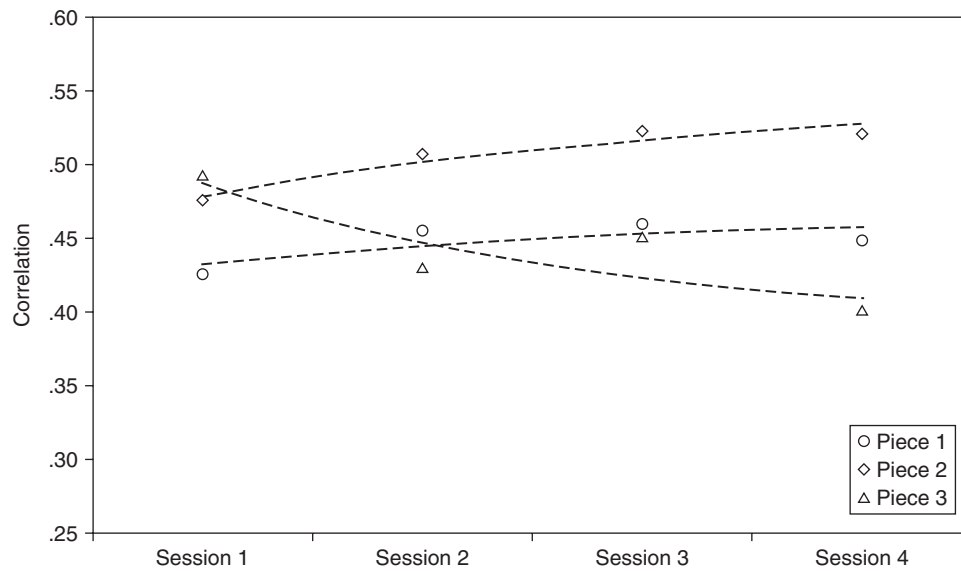


FIGURE 11. Average of correlations between listeners' movements and J7 for *reliable* performances over the four sessions.

Repeated Measures for unreliable movements revealed no significant effect over session. Figure 11 shows the results for reliable performances. The average number of significant listener-player correlations is between 40% and 55% of the total amount of reliable performances. For fragment P1 and P2, there was a trend towards an increasing amount of correlations over sessions, whereas for fragment P3, the trend decreased. GLM for Repeated Measures revealed a significant effect over piece ($p < .001$) (Table 4a) and a

significant effect over session ($p = .005$) (Table 4b). This result corresponds to previous findings that the movements depend on the nature of the musical pieces.

The number of correlations, still up to about 40% for P1 in the case of unreliable performances (Figure 10), may be explained by the fact that these performances may still contain movements that are in tune with those of the player. Indeed, unreliability is merely related to the fact that the listener did not repeat the movement in

TABLE 4. GLM for Repeated Measures Applied to the Data of Figure 11.

(A) Test for Differences Among Pieces

Source	Type IV Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	<i>p</i>	Partial Eta Squared
Intercept	67294.68	1	67294.68	386.32	.000	.82
Piece	3522.34	2	1761.17	10.11	.000	.19
Error (session)	15154.98	87	174.20			

(B) Test for Differences Among Sessions

Source	Type IV Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	<i>p</i>	Partial Eta Squared
Session	1.50	1	1.50	8.14	.005	.09
Session* Piece	0.33	2	0.17	0.91	.41	.02
Error (session)	16.06	87	0.19			

the subsequent trial. However, it is not excluded that one out of the two movement velocities of the repetitive trial has indeed some correspondence with the movement velocities of the player. The distinction between reliable and unreliable performances is indeed rather robust. However, in general, the percentages of the unreliable performances are significantly lower than the ones of the reliable performances. These results contribute to the hypothesis that the movement velocities of the listeners' arm tend to correlate with the movement velocities of the player's shoulder.

Discussion

Methodology

From a methodological point of view, we believe that the repeated listening design (using musical excerpts of 30 s) provides an interesting method for determining whether a music-driven movement is reliable or not. It was assumed that listeners who found a proper motor strategy to express their movement *in tune* with music (during these 30 seconds) would not radically change their strategy during 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 give a different performance during the second trial. Thus, high correlation between successive performances (e.g., P1—P1) would indicate *reliable performance*, while low correlation would indicate *unreliable performance*. The measurement of the correlation between the two repetitive performances thus offers an objective criterion for accessing movement reliability. The analysis of the movement data further reveals that better results can be obtained with warped movement patterns than with unwarped movement patterns. Improved results with warped data account for the fact that participants may show different adaptations (reacting faster or slower) to the presented musical excerpts.

Data about the listeners' subjective evaluation of their own performance (based on a questionnaire after each session, but not reported in detail in the present paper) are consistent with the objective observation of reliable performances. The participants were asked to evaluate the degree of difficulty, discrepancy, and satisfaction of their performances over the pieces and the sessions. Using cross tabulation (chi-square) it was found that the evaluation changed primarily over the pieces and slightly over the sessions. Although further details go beyond the scope of the present paper, it is of

interest to mention that, for reliable listener performances, a significant correlation was found between an objective measure (based on the repeated listening design explained above) and a subjective measure (based on the self-assessment of the listener's own performance). No such relationship was found for the unreliable performances.

Interlistener and Listener-Player Correlations

The results for interlistener correlations support the hypothesis that listeners share the movement velocity patterns of music-driven corporeal expressions, and thus the embodied perception of expression during listening to music. In addition, the results for listener-player correlations support the hypothesis that there may be a link between what listeners perceive as expression, and what the player produces as expression. This link is established through corporeal articulations that accompany music playing and music listening. The data seem to support the idea that listeners can somehow decode the expression that the player has encoded in the musical audio. The results also contribute to the hypothesis that the embodiment of music expression has a grounding in social communication.

Obviously, the two hypotheses, namely interlistener correlation, and listener-player correlation, are not completely independent. If the movement velocity patterns of the listeners correspond with the movement velocity patterns of the player ('s shoulder), then it can be expected that the movement velocity patterns of the listeners are also related with each other. Our analysis seems to confirm this relationship between the two hypotheses.

The Role of Music

The number of interlistener and listener-player correlations could depend on the nature of the musical fragment. The fragments with a melodic pitch line (P1 and P2) scored better than the fragment with a broken melodic, and thus more declamatory, pitch line (P3). It was noticed that fragment P3 displays an interesting difference between intra and intersubject analysis. The intrasubject analysis suggests that listeners tend to improve their embodied perception for P3 over session, while the intersubject 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 their embodiment solution differs from each other. In this particular example, the lack of

a shared intersubjective embodied listening solution may be due to the narrative character of the piece. The results suggest that melodic fragments may induce a higher degree of intersubjective share among listeners than narrative fragments, as the latter have more appeal to individual interpretation. In other words, listeners are more consistent among each other in the interpretation of melodic fragments than in the interpretation of narrative fragments, and this increases with session. This indicates that the listeners' shared music-driven corporeal articulations depend on the character of the music, that is, on musical parameters such as melodic line and prosody that have the power to activate movement.

Similar results have been found in De Bruyn et al. (2008). Although in this study, the group condition outperformed the individual condition, there was a major effect of musical piece, in the sense that some pieces, although similar in tempo, activated body movement better than other pieces. We believe that the differences between P1, P2, and P3 can be attributed to a similar effect of musical piece. The melodic lines of P1 and P2 are easier to express, and hence, their expression is easier to share. However, the precise relationship between acoustical features and musical structure on the one hand, and the movement features of its coarticulated corporeal expressions on the other hand needs further investigation.

The Role of Learning

The present study did not focus on learning, although the use of four sessions, mainly intended to test the consistency among the listeners' movements, revealed some interesting aspects. The significant improvement of the results over the four sessions for P1 and P2 seems to indicate that aspects of learning may be involved. The more the participants got acquainted with the guqin music, the more they shared each others' velocity of coarticulated corporeal expressions. This is in line with the idea that embodied listening responses may involve implicit learning, and hence, that the shared expression among participants may be enhanced in response to repeated exposure to music. However, learning may depend strongly on the character of the music, since P3 shows no such learning effect.

The Role of Shoulders

Analysis of the relationship between the movement velocity patterns of the joints of the player (Desmet & Leman, 2008) reveals that the head is often in advance

of the actual playing gesture. Studies on the kinematics of human movement of arms and hands show that the elbow and wrist are loosely coupled, whereas the shoulder and elbow are tightly coupled (Bosga, Meulenbroek, & Swinnen, 2003; d'Avella, Fernandez, Portone, & Lacquaniti, 2008; Lacquaniti and Soechting, 1982). Large-scale movements are carried out primarily using the proximal joints of the upper limb (shoulder and elbow), whereas small-scale movements are carried out by the most distal joints (wrist and fingers). Playing the guqin can be seen as the combination of both large-scale (shoulder—elbow) and small-scale (wrist—finger) movements. While playing, the head seems to indicate movement intention, and the movement of the shoulder can be considered as a large-scale movement that is related to the small-scale control of sound production gestures. The left arm is involved in large-scale transversal (or horizontal from the viewpoint of the player) movements from one position to another position on the guqin strings, and in small-scale (also transversal) movements that focus on pitch sliding. The large-scale movements are technical and not directly related to the sonic image, whereas the small-scale movements leave a direct sonic trace in terms of pitch modulation (such as glissandi and vibrato, as shown in Li & Leman, 2007). The right arm is mainly moving in the sagittal plane, which is perpendicular to the string (or vertical from the viewpoint of the player), for plucking purposes.

An analysis of the player's movements (Desmet & Leman, 2008) shows that the movement of the right shoulder (J7) and the right elbow (J8) correlate better (mean correlation coefficient for the three pieces = .61) than the movement of the left shoulder (J2) and the left elbow (J3) (mean correlation coefficient for the three pieces = .25). This is consistent with the finding that the listeners' movements correlate more with the right elbow than with the left elbow (see Figure 9), which seems to imply that the listeners' movements share patterns with the player's large-scale movements of expressive plucking. The fact that the arm and hand movements of the listeners are constrained by holding and moving a stick may constrain small-scale movements. However, movements of the wrist are possible, similar to conducting movements.

We further believe that the movement of the player's shoulder has an intentional component, though its precise goal may be vague. This hypothesis is based on the fact that plucking (by the right hand) provides energy to the string, which is then used for pitch sliding (by the left hand). However, during plucking, the player's attention is often focused on the pitch sliding because that action requires very precise and detailed movements.

Hence, the shoulder articulation that goes together with plucking may well be an expression of the movement pattern that has to be executed in a later stage after plucking. In that sense, the movement of the shoulder has an intentional character, rather than an executive character. Since listeners tend to move in correlation with the player's shoulder, we believe that it is justified to say that the shared expressive patterns among listeners and player contain an aspect of intentionality, although the precise goal of the movement may remain vague. Kant (1790/1924) already drew attention to the notion of *sensus communis* as the capacity for sensing goal-directeness without goal. He conceived it as the core capacity for aesthetic appreciation, which human participants share among each other.

The above findings seem to support the hypothesis (i) that listeners can decode to a certain degree the expression of the musical stimulus through corporeal articulations, and (ii), that, in some cases, these corporeal articulations may reflect the expressive articulations of the performer.

General Discussion

The ability to decode the expressive code from music through embodiment can be linked with a social component and a language component.

Indeed, one of the major effects known in social anthropology is that music contributes to the development of a personal self of young people, and that it fosters social bonding and group identity (Freeman, 2000; Gregory, 1997; Hargreaves & North, 1997). Although this social anthropological aspect has not been addressed explicitly in the present study, our methodology of monitoring embodiment may be highly relevant in view of studies that address social music interaction. Expressive social behavior often relies on nonconscious imitation or mimicry, which shows itself in movement synchrony, behavior matching, as effectuated in speech patterns, facial expressions, emotions, moods, postures, gestures, mannerisms, and idiosyncratic movements (Bargh & Chartrand, 1999; Lakin, Jefferis, Cheng, & Chartrand, 2003; Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005). Within a social group setting, a participant is more likely to get along harmoniously with others in the group if it is behaving similarly to them, compared with being "out of sync" and behaving differently. Also in the visual domain, observers are most sensitive to their own movements, compared to those of friends, and then those of strangers (Loula, Prasad, Harber, & Shiffrar, 2005). In that context, music may be considered a particular case in which expressive behaviors are driven by a sonic

source. Our study suggests that these behaviors may be partly shared, and that embodiment forms the core of this *sensus communis* (Kant, 1790/1924) for music expression. Our study suggests that music, through embodiment, can be a strong driver of social interaction. This finding is in line with recent studies on empathy and social interaction, in which the 'mirror system' plays a central role (see e.g., Gallese, 2003; Knoblich & Sebanz, 2006).

As to language, McNeill (2005) and Gallagher (2005) have argued that hand gestures that accompany speech can be seen as a motor component of speech, necessary to create the narrative space that is shared in a communicative situation (see also Brown et al., 2006; Sparing et al., 2007). In this approach, expressive (co)articulations of the hand are not just seen as a supplement of language, but as a component that is fully integrated with language. We believe that the result of the present study is in agreement with the hypothesis that spoken language and music have a common motor component that shows itself in expressive gesture. In both music and language, it is likely that embodiment creates a shared space for the communication of expression.

Our study suggests that the communication of an expressive code through music forms part of a mechanism where listeners, in order to bridge the semantic gap between sonic cues and meaning, embody music and thus engender a plausible or virtual action to music based on their own action repertoire (Leman, 2007). In line with the embodied music cognition paradigm, this embodiment is based on a mirror system that is rooted in a shared neuronal basis for perception and action. Through embodied listening, music perception can be said to be grounded in the listener's subjective action repertoire and its perceived expression can be shared and signified by a community.

Conclusion

The present study developed a methodology to study the hypothesis that expressive patterns in music can be generated, perceived, and shared among player and listeners. The methodology is based on the measurement of movements, using a design in which listeners have to repeat their movement patterns in order to be sure that their coarticulations were regular or reliable.

The epistemological background for this study is rooted in the paradigm of embodied music cognition, which assumes a close relationship between perception and action through a mirroring mechanism. In this paradigm, the human body is seen as a natural mediator between subjective experience and physical reality. Corporeal coarticulations of music playing and music

listening can be seen as expressions of this natural mediation.

The comparison of the movement velocity patterns suggest that listeners share their sensitivity for musical expression and that they are able to display this sensitivity through music-driven corporeal articulations. The study also revealed that the movement velocity patterns of the player's shoulder correlates with the movement velocity patterns of the listeners' arms, which suggests that their corporeal coarticulations are shared to a certain

degree. Further analysis of the shoulder patterns let us believe that the sharing of musical expression has an intentional character.

Author Note

Correspondence concerning this article should be addressed to Marc Leman, Department of Musicology, Ghent University, Blandijnberg 2, B-9000, Ghent, Belgium. E-MAIL: Marc.Leman@UGent.be

References

- ARISTOTLE, & MCKEON, R. P. (2001). *The basic works of Aristotle*. New York: Modern Library.
- BARGH, J. A., & CHARTRAND, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*, 462-479.
- BECKING, G. (1928). *Der musikalische Rhythmus als Erkenntnisquelle* [Musical rhythm as source of knowledge]. Augsburg: B. Filser.
- BERTHOZ, A. (2000). *The brain's sense of movement*. Cambridge, MA: Harvard University Press.
- BOSGA, J., MEULENBROEK, R. G. J., & SWINNEN, S. P. (2003). Stability of inter-joint coordination during circle drawing: Effects of shoulder-joint articular properties. *Human Movement Science*, *22*, 297-320.
- BROWN, S., MARTINEZ, M. J., & PARSONS, L. M. (2006). Music and language side by side in the brain: A PET study of the generation of melodies and sentences. *European Journal of Neuroscience*, *23*, 2791-2803.
- CAMURRI, A., VOLPE, G., DE POLI, G., & LEMAN, M. (2005). Communicating expressiveness and affect in multimodal interactive systems. *IEEE Multimedia*, *12*, 43-53.
- CHEN, J. L., ZATORRE, R. J., & PENHUNE, V. B. (2006). Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *Neuroimage*, *32*, 1771-1781.
- CLAYTON, M. R. L. (2007). Observing entrainment in music performance: Video-based observational analysis of Indian musicians' tanpura playing and beat marking. *Musicae Scientiae*, *11*, 27-59.
- CLYNES, M. (1977). *Sentics: The touch of emotions* (1st ed.). Garden City, NY: Anchor Press.
- CUMMING, N. (2000). *The sonic self: Musical subjectivity and signification*. Bloomington, IN: Indiana University Press.
- D'AUSILIO, A., ALTENMÜLLER, E., BELARDINELLI, M. O., & LOTZE, M. (2006). Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece. *European Journal of Neuroscience*, *24*, 955-958.
- D'AVELLA, A., FERNANDEZ, L., PORTONE, A., & LACQUANITI, F. (2008). Modulation of phasic and tonic muscle synergies with reaching direction and speed. *Journal of Neurophysiology*, *100*, 1433-1454.
- DE BRUYN, L., LEMAN, M., DEMEY, M., & MOELANTS, D. (in press). Does social interaction activate music listeners? In R. Kronland-Martinet, S. Ystad, & K. Jensen (Eds.), *Computer music, modeling and retrieval. Genesis of meaning of sound and music. Lecture notes in computer science*. Berlin, Heidelberg: Springer-Verlag.
- DE BRUYN, L., LEMAN, M., & MOELANTS, D. (2008, August). *Quantifying children's embodiment of musical rhythm in individual and group settings*. Paper presented at the meeting of the International Conference of Music Perception and Cognition, Sapporo, Japan.
- DE POLI, G. (2004). Methodologies for expressiveness modelling of and for music performance. *Journal of New Music Research*, *33*, 189-202.
- DESMET, F., & LEMAN, M. (2008). *The coordination of joints during guqin playing*. Manuscript in preparation.
- DESMET, F., LEMAN, M., LESAFFRE, M., & DE BRUYN, L. (in press). Statistical analysis of human body movement and group interactions in response to music. In H-H Bock et al. (Eds.), *Studies in classification, data analysis and knowledge organization*. Berlin, Heidelberg: Springer-Verlag.
- FREEMAN, W. (2000). A neurobiological role of music in social bonding. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 411-424). Cambridge, MA: MIT Press.
- FRIBERG, A., & SUNDBERG, J. (1999). Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *Journal of the Acoustical Society of America*, *105*, 1469-1484.
- FRIBERG, A., SUNDBERG, J., & FRYDEN, L. (2000). Music from motion: Sound level envelopes of tones expressing human locomotion. *Journal of New Music Research*, *29*, 199-210.
- GABRIELSSON, A. (2003). Music performance research at the millennium. *Psychology of Music*, *31*, 221-272.
- GALLAGHER, S. (2005). *How the body shapes the mind*. Oxford: Clarendon Press.
- GALLESE, V. (2003). The manifold nature of interpersonal relations: The quest for a common mechanism. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, *358*, 517-528.

- GREGORY, A. (1997). The roles of music in society: The ethnomusicological perspective. In D. J. Hargreaves & A. C. North (Eds.), *The social psychology of music* (pp. 123-140). Oxford: Oxford University Press.
- HANSLICK, E. (1891). *Vom musikalisch-schönen ein Beitrag zur Revision der Ästhetik der Tonkunst* [The beautiful in music: A contribution to the revisal of musical aesthetics]. Leipzig: Barth.
- HARGREAVES, D. J., & NORTH, A. C. (Eds.). (1997). *The social psychology of music*. Oxford: Oxford University Press.
- HATTEN, R. S. (2003). Thematic gestures, topics, and tropes. In E. Tarasti (Ed.), *Musical semiotics revisited* (pp. 80-91). Helsinki: Hakapaino.
- HOMMEL, B., MUSSELER, J., ASCHERSLEBEN, G., & PRINZ, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24, 849-878.
- HUME, D., SELBY-BIGGE, L. A., & NIDDITCH, P. H. (1975). *Enquiries concerning human understanding and concerning the principles of morals* (3d ed.). Oxford: Clarendon Press.
- JEANNEROD, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, 17, 187-202.
- JUSLIN, P. N., & LAUKKA, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, 33, 217-238.
- KANT, I. (1790/1924). *Kritik der Urteilskraft* [Critique of judgement] (6th ed.). Leipzig: F. Meiner.
- KELLER, P. (2008). Joint action in music performance. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting intersubjectivity: A cognitive and social perspective on the study of interaction* (pp. 205-221). Amsterdam: IOS.
- KELLER, P., KNOBLICH, G., & REPP, B. (2006). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and Cognition*, 16, 102-111.
- KNOBLICH, G., & SEBANZ, N. (2006). The social nature of perception and action. *Current Directions in Psychological Science*, 15, 99-104.
- LACQUANITI, F., & SOECHTING, J. F. (1982). Coordination of arm and wrist motion during a reaching task. *Journal of Neuroscience*, 2, 399-408.
- LAKIN, J. L., JEFFERIS, V. E., CHENG, C. M., & CHARTRAND, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, 27, 145-162.
- LEMAN, M. (2007). *Embodied music cognition and mediation technology*. Cambridge, MA: MIT Press.
- LEMAN, M., & CAMURRI, A. (2005). Understanding musical expressiveness using interactive multimedia platforms. *Musicae Scientiae, Special issue 2005/06*, 209-233.
- LESAFFRE, M., DE VOOGDT, L., LEMAN, M., DE BAETS, B., DE MEYER, H., & MARTENS, J. P. (2008). How potential users of music search and retrieval systems describe the semantic quality of music. *Journal of the American Society for Information Science and Technology*, 59, 695-707.
- LI, H., & LEMAN, M. (2007). A gesture-based typology of sliding-tones in guqin music. *Journal of New Music Research*, 36, 61-82.
- LIPPS, T. (1903). *Ästhetik: Psychologie des Schönen und der Kunst* [Aesthetics: Psychology of beauty in art]. Hamburg und Leipzig: L. Voss.
- LOULA, F., PRASAD, S., HARBER, K., & SHIFFRAN, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology-Human Perception and Performance*, 31, 210-220.
- MCNEILL, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- MION, L., & DE POLI, G. (2008). Score-independent audio features for description of music expression. *IEEE Transactions on Audio Speech and Language Processing*, 16, 458-466.
- NETTHEIM, N. (1996). How musical rhythm reveals human attitudes: Gustav Becking's theory. *International Review of the Aesthetics and Sociology of Music*, 27, 101-122.
- NIEDENTHAL, P. M., BARSALOU, L. W., WINKIELMAN, P., KRAUTH-GRUBER, S., & RIC, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, 9, 184-211.
- PALMER, C. (2005). Time course of retrieval and movement preparation in music performance. *Annals of the New York Academy of Sciences*, 1060, 360-367.
- PHILLIPS-SILVER, J., & TRAINOR, L. J. (2008). Vestibular influence on auditory metrical interpretation. *Brain and Cognition*, 67, 94-102.
- REPP, B. H. (1993). Music as motion: A synopsis of Alexander Truslit's (1938) *Gestaltung und Bewegung in der Music*. *Psychology of Music*, 12, 48-72.
- REPP, B. H., & KNOBLICH, G. (2004). Perceiving action identity—How pianists recognize their own performances. *Psychological Science*, 15, 604-609.
- SAKOE, H., & CHIBA, S. (1978). Dynamic-programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics Speech and Signal Processing*, 26, 43-49.
- SPARING, R., MEISTER, I. G., WIENEMANN, M., BUELTE, D., STAEDTGEN, M., & BOROOJERDI, B. (2007). Task-dependent modulation of functional connectivity between hand motor cortices and neuronal networks underlying language and music: A transcranial magnetic stimulation study in humans. *European Journal of Neuroscience*, 25, 319-323.
- STYNS, F., VAN NOORDEN, L., MOELANTS, D., & LEMAN, M. (2007). Walking on music. *Human Movement Science*, 26, 769-785.

SUNDBERG, J. (ED.). (2003). *Research in music performance* (Special issue of *Journal of New Music Research*). Lisse: Swets & Zeitlinger.

TRUSLIT, A. (1938). *Gestaltung und Bewegung in der Musik* [Figuration and motion in music]. Berlin-Lichterfelde: C.F. Vieweg.

WANDERLEY, M. M., VINES, B. W., MIDDLETON, N., MCKAY, C., & HATCH, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, 34, 97-113.

WIDMER, G., & GOEBL, W. (2004). Computational models of expressive music performance: The state of the art. *Journal of New Music Research*, 33, 203-216.

WILSON, M., & KNOBLICH, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131, 460-473.

ZATORRE, R. J., CHEN, J. L., & PENHUNE, V. B. (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8, 547-558.

Appendix

Dynamic time warping (DTW) is a time series alignment algorithm that was originally developed for speech recognition (Sakoe & Chiba, 1978). The goal was to align two time series by warping the time axis iteratively until an optimal match between the two sequences was found.

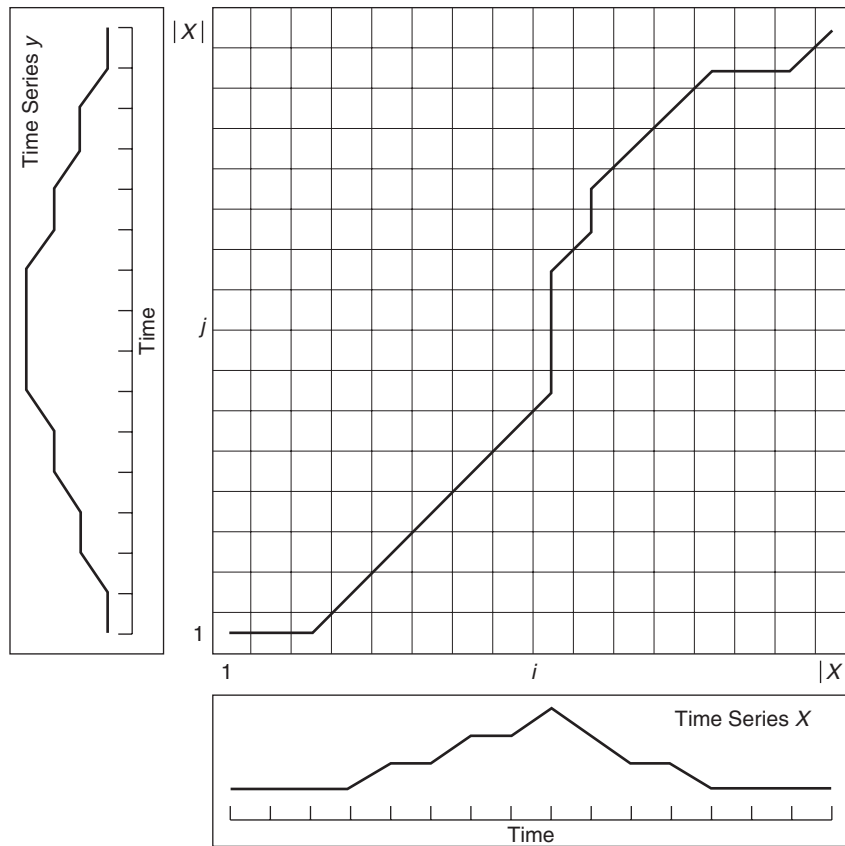
The technique can be described as follows:

Given two time series:

$$A = x_1, x_2, \dots, x_i \dots, x_n$$

$$B = y_1, y_2, \dots, y_i \dots, y_m$$

The two series are arranged on the sides of a grid, (one on the top and the up the left hand side) both starting at the bottom left of the grid.

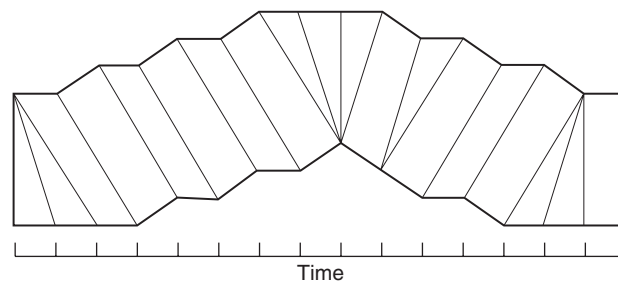


Inside each cell the Euclidean distance is calculated for the corresponding x,y pair. To find the best match or alignment between the series a path has to be found through the grid which minimizes the total distance between the two series. The procedure for computing this overall distance involves finding all possible routes through the grid and for each one compute the overall distance.

Using optimizations and constraints reduces the number of possible paths and reduces the calculation time. The following constraints were used in this analysis:

- **Monotonic condition:** the path will not turn back on itself, both the i and j indexes either stay the same or increase, they never decrease.
- **Continuity condition:** the path advances one step at a time. Both i and j can only increase by, at most, one on each step along the path.
- **Boundary condition:** the path starts at the bottom left and ends at the top right.
- **Warping window condition:** a good path is unlikely to wander very far from the diagonal. The distance that the path is allowed to wander is the window width. The maximum width was chosen to be three.
- **Slope constraint condition:** The path should not be too steep or too shallow. This prevents short sequences matching too long ones. The condition is expressed as a ratio p/q where p is the number of steps allowed in the same (horizontal or vertical) direction. After p steps in the same direction is not allowed to step further in the same direction before stepping at least q times in the diagonal direction.

The power of the DTW algorithm is that instead of finding all possible routes through the grid that satisfy the above conditions, the DTW algorithm works by keeping track of the cost of the best path to each point in the grid. During the calculation process of the DTW grid it is not known which path is minimum overall distance path, but this can be traced back when the end point is reached.



APPENDIX 2.

It is well known that one of the major disadvantages of DTW is due to the limitations of local disturbances in the series. This results in less reliable outcomes especially if the original series has a greater proportion of noise and hence a less pronounced pattern. A lower correlation of the warped series with respect to the original (Euclidean) correlation is an indicator for a low signal to noise ratio. In the guqin analysis this was found for Piece 3. Since the development of DTW (in the 1970s) other techniques, such as hidden Markov modeling and Neural Networks, have been developed to cope with this drawback of DTW.



FIGURE 1 (LEMAN ET AL.). Guqin player and motion capturing system.



FIGURE 2 (LEMAN ET AL.). Recording of embodied music listening via joystick.

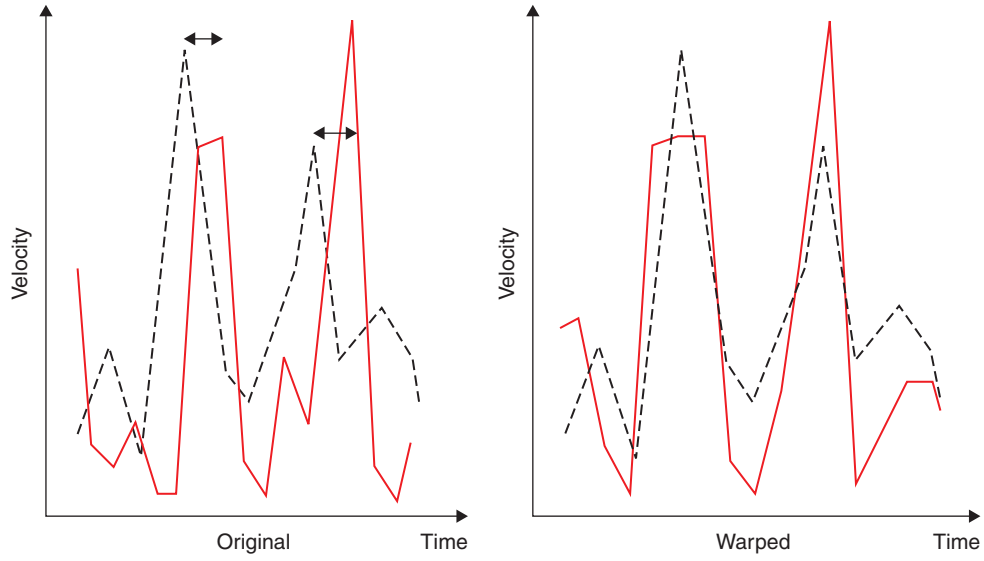


FIGURE 3 (LEMAN ET AL.). The effect of dynamic time warping. (a) Comparison between unwarped (solid line) and reference (dotted line), (b) comparison between warped (solid line) and reference (dotted line).

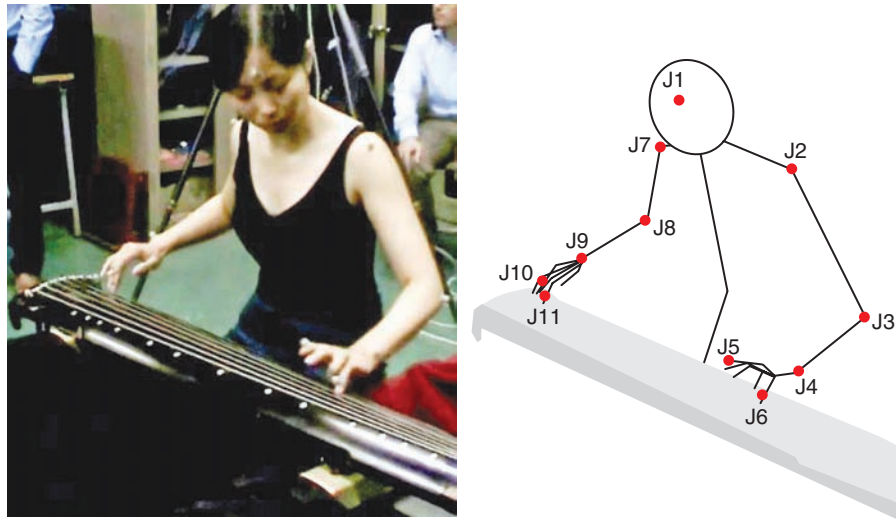


FIGURE 8 (LEMAN ET AL.). Picture of the player and related stick figure with joints labeled (see text).