

From expressive gesture to sound

The development of an embodied mapping trajectory inside a musical interface

Pieter-Jan Maes · Marc Leman · Micheline Lesaffre · Michiel Demey · Dirk Moelants

Received: date / Accepted: date

Abstract This paper contributes to the development of a multimodal, musical tool that extends the natural action range of the human body to communicate expressiveness into the virtual music domain. The core of this musical tool consists of a low cost, highly functional computational model developed upon the Max/MSP platform that (1) captures real-time movement of the human body into a 3D coordinate system on the basis of the orientation output of any type of inertial sensor system that is OSC-compatible, (2) extract low-level movement features that specify the amount of contraction/expansion as a measure of how a subject uses the surrounding space, (3) recognizes these movement features as being expressive gestures, and (4) creates a mapping trajectory between these expressive gestures and the sound synthesis process of adding harmonic related voices on an in origin monophonic voice. The concern for a user-oriented and intuitive mapping strategy was thereby of central importance. This was achieved by conducting an empirical experiment based on theoretical concepts from the embodied music cognition paradigm. Based on empirical evidence, this

paper proposes a mapping trajectory that facilitates the interaction between a musician and his instrument, the artistic collaboration between (multimedia) artists and the communication of expressiveness in a social, musical context.

Keywords multimodal interface · mapping · inertial sensing technique · usability testing

1 Introduction

1.1 Theoretical framework

Playing music requires the control of a multimodal interface, namely, the music instrument, that mediates the transformation of bio-mechanical energy to sound energy, using feedback loops based on different sensing channels, such as auditory, visual, haptic and tactile channels. In recent decennia, a lot of attention has been devoted to the development of electronic multimodal interfaces for music [4,6,9]. The basic problem of these interfaces concerns the fact that the mediation between the different modalities (basically from movement to sound) has an arbitrary component, which is due to the fact that the energies of the modalities are transformed into electronic signals. This is in contrast with traditional instruments, where energetic modalities are mechanically mediated and where the user gets a natural feeling of the causality of the multimodal interface.

So far, research has focused on theoretical reflections on the possible connection between movement and sound [41,5,1,13,23] and especially also on the development of all kinds of electronic multimodal interfacing technologies, including models and practical designs for the mapping of performed action to sound synthesis parameters [35,21,2,38,20]. However, relatively less attention has been devoted to the idea of an empirical solution to the mapping problem. Such a solution would be based on experiments that probe

P. -J. Maes, M. Leman, M. Lesaffre, M. Demey, D. Moelants
IPEM - Dept. of Musicology, Ghent University, Blandijnberg 2,
B-9000, Ghent Tel.: +32 (0)9 264 4126
Fax: +32 (0)9 264 4143
E-mail: pieterjan.maes@UGent.be

M. Leman
E-mail: marc.leman@UGent.be

M. Lesaffre
E-mail: micheline.lesaffre@UGent.be

M. Demey
E-mail: michiel.demey@UGent.be

D. Moelants
E-mail: dirk.moelants@UGent.be

the natural tendencies for multimodal mappings. The theoretical grounding for such approach is found in the embodied music cognition paradigm, in which the importance of gesture and corporeal imitation as a basis of understanding musical expressiveness is stressed [27, 28]. In this view, multimodal interfaces for music are approached in terms of mediation technologies that connect with action-relevant cues (i.e. affordances). Musical intentionality and expressiveness can be attributed by a mirroring process that relates these cues to the subjects own action-oriented ontology.

The present paper aims at identifying mapping strategies for multimodal music interfaces, using these concepts of embodiment, action-perception coupling and action-oriented ontology as a starting point. The goal is to better understand bodily movement in relation to sound from the viewpoint of the peripersonal space, that is, the space immediately surrounding a person's body which can be reached by the limbs [40, 19]. This study will focus on the relation between the dynamically changing pattern of the upper limbs in terms of contraction and expansion and the communication of musical expressiveness. The choice for this particular movement feature is made in line with findings of previous research indicating that: (I) music intuitively stimulates movement in the peripersonal space [34, 17, 18, 12], (II) the human body is an important channel of affective communication [15, 16, 11, 3, 37], (III) the upper body features are most significant in conveying emotion and expressiveness [24], (IV) the movement size and openness can be related to the emotional intensity of the musical sound production [14, 7], and (V) an open body position in contrast to a closed body position reinforces the communicators intent to persuade [31, 30]. It is assumed that a better understanding of the connection between this spatio-kinetic movement feature and expressive features in relation to multimodal interfaces may provide a cue to the solution of the mapping problem in a number of application contexts.

1.2 Methodological framework

The methodology of this paper will follow the layered, conceptual framework proposed by Camurri [6, 8, 9, 27]. This model starts with modelling movement on a pure physical level, followed by the extraction of features that are subsequently mapped into gestural trajectories and linked with high-level structures and concepts. This framework makes it possible to establish connections between the sensory, the gestural and the semantic level.

The first part of this paper focuses on the development of a low cost, highly functional Max/MSP algorithm that enables measurement and modelling of human movement on the basis of the orientation output of inertial sensing systems. The platform is innovative in the way it supports basically every sensor that outputs orientation data. The second

part of this paper integrates this modelling platform into an empirical-experimental framework. An experiment is conducted in order to study the natural, corporeal resonance behaviour of subjects as a response to a specific musical feature integrated in four different pre-recorded sound stimuli. The musical feature that is under investigation is termed with *one-to-many alternation*. It concerns the musical process of adding and subsequently removing extra harmonic related voices on an in origin monophonic voice. As a result, this study wants to put in evidence the effect of the contrast between a solo voice and multiple harmonic voices on bodily movement behaviour. After pure physical measurements of human movement, low-level movement features are extracted. Based on the effort/shape theory of Laban [26] a particular interest lies on spatio-kinetic movement features that measure the amount of contraction/extension of the upper limbs in the peripersonal space or kinesphere of subjects. This feature will be defined by taking into account the distance between (1) the elbows relative to each other and (2) the wrists relative to each other. Then, the extracted movement features are investigated in relation to the auditory feature and mapped into gesture trajectories that describe the one-to-many alternation. A next layer defines how these specific, spatio-kinetic gestural trajectories can be associated with an expressive content forming *expressive gestures* [25, 4, 8]. This will be realised by integrating verbal descriptions related to emotion and expression into a model for semantic description of music [29]. The gestural trajectory may then be connected with parameters that control the synthesis of in principal every kind of energetic modality [27]. However, we will limit these possibilities in this paper to the proposal of a system that enables the recognition and extraction of expressive gestures that are subsequently mapped to control a sound process that corresponds with the one-to-many alternation.

2 Technical Setup - Motion Sensing and 3D position determination

Inertial sensors were used to measure the movement of the upper body of the user. The sensors enable a mobile setup, which is often useful in an artistic context. There is no problem of visual occlusion of marker points or lighting conditions, although the large mass and shape of the sensors can be problematic. Inertial sensors do not provide an absolute 3D position but it is possible to determinate the relative 3D position of the joints of the upper body with respect to a fixed point on the body, using only the orientation output of the sensors together with fixed lengths of the different body parts. In what follows the motion sensors are described in more detail, together with the software used and the different steps in the algorithm that determines the relative 3D position of the upper body.

In this study, five commercial inertial sensors are used from Xsens (MTx XBus Kit). They are positioned on the chest, the upper arms and the lower arms as shown in figure 1 (above). Flexible straps with Velcro are used to attach the sensors to the body. The sensors are daisy chained with wires and connected to a battery powered central unit (XBus) that is attached to the subjects hip with a belt. From this central unit the data is transmitted wirelessly over a Bluetooth connection. This setup enables a sampling rate of 25Hz of the quaternion information of the five motion sensors. This data is collected on a MacBook Pro laptop running a standalone program that collects the data from the Bluetooth port and converts this into the OSC protocol. The sensor data is then send to a Max/MSP program that calculates the relative 3D position of the four joints (the elbows and wrists) of the upper body. A visualization, generated in Jitter, is shown in figure 1 (below).

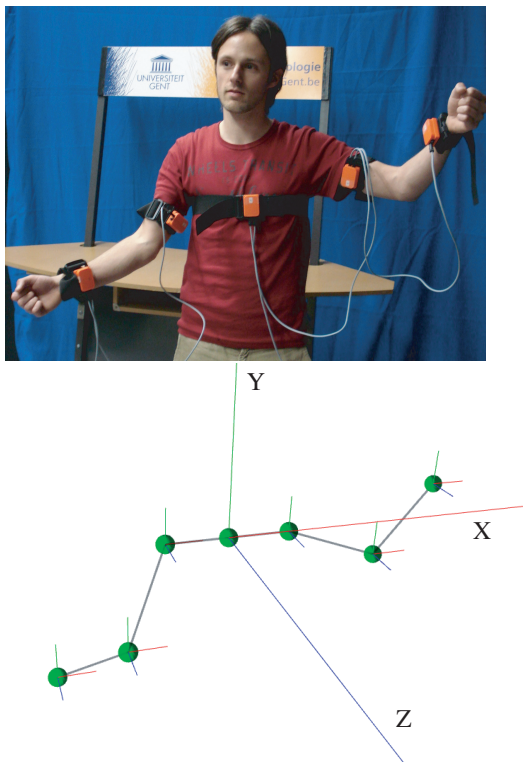


Fig. 1 Above: person equipped with five inertial MTx sensors from Xsens on the upper body. Below: a stick-figure representation of the upper body generated in Jitter (Cycling '74).

To calibrate the system the subject had to stand up with horizontally stretched arms, so that a T-shape is formed with the torso and the arms. In this posture all sensors are orientated in the same way, which allows a calibration of the system. The 3D coordinate system attached to the body of the subject is called the local system. Its vertical direction is defined as the Y-axis, the horizontal direction as the X-axis pointing to the left from the viewpoint of the user and

coinciding with the arms when the T-pose is assumed and the Z-axis is defined as pointing forward in the horizontal plane. As such, a right-handed coordinate system is defined. The origin of this system is located on the torso in the middle of the line connecting the shoulders. The orientation of this local coordinate system is fixed by the orientation of the motion sensor attached to the chest.

The position of the shoulders is then fixed on the X-axis at a distance of 17 cm from the origin, resulting in coordinates $(-0.17, 0.0, 0.0)$ and $(0.17, 0.0, 0.0)$, where the units are expressed in meters. The length of the upper arms and lower arms are fixed to 26 cm and 27 cm respectively. To obtain the position of the elbows one has to calculate the relative orientation of the upper arm with respect to that of the chest. This is obtained through the Hamiltonian product of the quaternions of the chest (q_c) and that of the upper arm (q_u) where the first quaternion is conjugated:

$$q = q_c^* \cdot q_u \quad (1)$$

The resulting quaternion is converted into three Euler angles namely the pitch (rotation around the Y-axis), roll (rotation around the X-axis) and yaw (rotation around the Z-axis). When taking the biomechanical constraints of the shoulder joint into account, one can see that there are only two rotations that influence the position of the elbow. These rotations are the flexion/extension (rotation in the horizontal plane around the vertical direction) and abduction/adduction (rotation in the vertical plane around the horizontal direction). Using the Euler angles, the length of the upper arm and the position of the shoulders the 3D position of the elbow in the local coordinate system can then be calculated.

Once the positions of the elbows are calculated, the position of the wrists can be obtained in the same way as described above, using the relative difference in orientation between the upper arm and the lower arm. For the case of the elbow joint there is only the flexion/extension rotation that has an effect on the position of the wrists. The results from this rotation is a vector \mathbf{w} describing the position of the wrist in reference to the elbow. A crucial step is the transformation of this vector \mathbf{w} from the frame oriented in the elbow to the local coordinate system with the orientation of the sensor on the chest. This transformation is accomplished through the use of the following formula:

$$\mathbf{w}' = q \cdot \mathbf{w} \cdot q^* \quad (2)$$

where q is the relative orientation of the upper arm q_u with respect to the chest q_c as defined in equation 1 and \mathbf{w}' is the resulting 3D position of the wrist in orientation of the local coordinate system. To make the implementation of formula 2 more clear one can write out the quaternions explicitly as follows:

$$\mathbf{w}' = (q_w, iq_x, jq_y, kq_z) \cdot (0, iw_x, jw_y, kw_z) \cdot (q_w, -iq_x, -jq_y, -kq_z) \quad (3)$$

By making use of the Hamiltonian product and the calculus of the quaternions one can obtain the following result:

$$w'_x = w_x(q_xq_x + q_wq_w - q_yq_y - q_zq_z) + w_y(2q_xq_y - 2q_wq_z) + w_z(2q_xq_z + 2q_wq_y) \quad (4)$$

$$w'_y = w_x(2q_wq_z + 2q_xq_y) + w_y(q_wq_w - q_xq_x + q_yq_y - q_zq_z) + w_z(-2q_wq_x + 2q_yq_z) \quad (5)$$

$$w'_z = w_x(-2q_wq_y + 2q_xq_z) + w_y(2q_wq_x + 2q_yq_z) + w_z(q_wq_w - q_xq_x - q_yq_y + q_zq_z). \quad (6)$$

When the resulting coordinates of vector \mathbf{w}' are added with the coordinates of the elbow, the coordinates of the wrist in the local frame are obtained. From this point on the 3D coordinates of the joints of the upper body are fully determined and can be used both in the visualization of an avatar and in the measurement of the expressive movement of the subject.

3 Set-up of the empirical-experimental framework

Like it was quoted in section 1 the experiment proposed in this paper starts from the embodied music cognition theory that states that musical involvement relies on motor imitation [27]. That is why in this experiment research is done concerning the motor resonance behaviour of subjects in response to the musical process called the one-to-many alternation (see section 1). The hypothesis of the experiment is that there exists a coupling between the auditory information received from the pre-recorded sound stimuli and the motor behaviour as a spontaneous re-action to this sensory input (cfr. sensorimotor coupling). Moreover, it is assumed that (1) the spatial characteristics of this motor reaction in terms of contraction and expansion of the upper limbs in peripersonal space shows communalities over the different subjects participating in the experiment and (2) the sensorimotor coupling arouses a similar expressive percept among the different subjects. If this is indeed the case, knowledge about this sensorimotor coupling could provide grounds for the development of a mapping trajectory facilitating an intuitive, expressive interaction between man, machine and social/artistic environment.

3.1 Subjects

25 subjects, aged 21-27 (mean: 23.2), 5 male, 20 female, participated in the experiment. 20 of them reported had at least seven years of musical education, four of the non-musicians and two musicians had between 1 and 10 years of dance education.

3.2 Auditory stimuli

Table 1 Characteristics of the four musical stimuli.

Stimuli	Timbre	Mode	Direction	Cycle length	Tempo	Repeats
A	Voice	Major	Below	12	90	5
B	Voice	Major	Below	4	90	10
C	Piano	Minor	Above	4	90	10
D	Piano	Minor	Above	4	120	14

The experiment presented in this study investigates the effect of a changing number of musical, harmonic voices on bodily movement. Therefore, four sound stimuli were made and recorded in advance of the actual experiment emphasizing this specific musical feature. They all exist of one single musical pattern repeated over time (see figure 2), characterised by the appearance and subsequently disappearance of extra harmonic voices upon an in origin single voice (a musical feature that was termed with one-to-many alternation). The extra voices are the harmonic third and fifth. The starting tone is always A4 (440 Hz). As a result, a single voice is gradually building up to a triad chord and subsequently fading away back to a single sounding voice. This effect of harmonization, applied in advance of the experiment, was done differently for the different stimuli. In the first two stimuli, discrete quarter notes produced by the human voice were used as input of a harmonizer (i.e. MuVoice) that electronically synthesized the extra voices by real-time pitch shifting. The volume of the two extra voices was controlled and recorded manually with the help of a USB controller (Evolution UC-16) in a way it resulted in the patterns defined in figure 2. In the last two stimuli, piano sounds were used. There, the harmonization of the discrete quarter notes was applied directly by a pianist by adding respectively the third and fifth. In order to find out if shared strategies of movement can be found using different stimulus material, pitch direction, mode, timbre, rhythm and tempo were varied. In order to avoid confusion with the effect of a rising or falling pitch on the movements, in two of the stimuli the triad was added above the starting tone, while in the other two it was added below the starting tone. Similarly, a possible effect of the mode was dealt with by using a major triad in the first two stimuli and a minor triad in the other two. The main characteristics of the four series are summarized in table 1, and a transcription in musical notation is given in figure 2. These show the variation in rhythm and tempo, and the use of two different timbres: a recording of a soprano voice singing A4 in the first two and the A4 played on a piano in the last two stimuli. The patterns given in figure 2, were repeated between 5 and 14 times, so the total length of one series was always between 45 and 60 seconds.



Fig. 2 Representation of the auditory stimuli by means of note patterns.

3.3 Procedure

The subjects were subsequently equipped with the five inertial sensors (Xsens, MTx XBus Kit) as explained in section 2 (see figure 1). The subjects were instructed to move along with the variation they noticed in the pre-recorded sound stimuli, using their arms and upper body. Neither details about the stimuli were given, nor more precise specifications on the way to move to be sure that the results arose from a natural and intuitive interaction between the subjects and the auditory stimuli.

Before starting the actual experiment, the participants performed a test trial. The researchers created and recorded the test stimulus in advance according to the same principle described in section 3.2. However, a sine wave was used instead of the human voice or piano sound. This test trial allowed the subjects to get acquainted with the task and the feeling of the sensors attached to their body. It also allowed to check the technical setup and eventually to fix some problems.

After the test trial, the subjects performed the same task while listening to stimulus A. Immediately after the measurement, they were asked to give a verbal, subjective description of how they experienced the evolution in the auditory stimuli using affective adjectives that they could choose spontaneously at own will. This procedure was repeated for stimuli B, C and D.

4 Results of the experiment

4.1 Analysis of the one-to-many variation in the auditory stimuli

The intent of this study is to quantify the relation between the musical structure inherent in the four pre-recorded sound stimuli and the bodily response of subjects in terms of the contraction/expansion pattern of the upper limbs in peripersonal space. The quantification of this relation requires a

comparison between the two features on a low-level, physical level. Therefore, the musical feature under investigation (i.e. one-to-many alternation) needs to be extracted from the complex acoustic signal representing each pre-recorded sound stimulus. The process to obtain this feature in physical format is explained in more detail in this section. It demands several operations whereby the complexity of the original acoustic signal is reduced in a high degree, yet without losing essential information. First, a pitch-tracking algorithm is executed on the acoustic signal of the four pre-recorded sound stimuli in order to obtain the frequencies inherent in the sound as a function of the time. By selection of (i.e. filtering out) the fundamental frequencies related to the three different voices (tonica, third, fifth), it could be observed how the presence (i.e. amplitude) of the multiple voices evolves over time and as such establish the effect of harmonization termed as one-to-many alternation. Second, a continuous contour is created that gives an indication of the evolving harmonization. This was done by adding the different amplitudes of the third and fifth and normalizing this result from 0 to 1. The minimum value of the signal means that only the root voice is present, the middle value means root and third are present and the maximum value means root, third and fifth are present. The continuous contour is obtained by applying an FFT-based filter to the data where the lower phazors were retained. The result (figure 3) is a low-level, physical signal specifying the musical feature providing a means to quantify the relation between sound and movement.

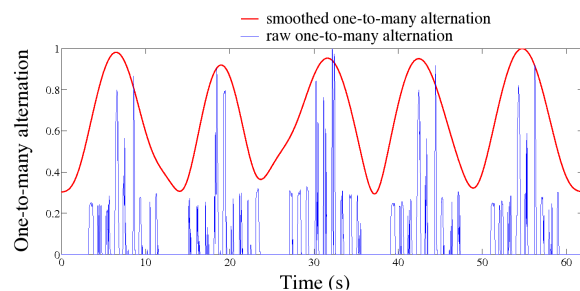


Fig. 3 Visualisation of the one-to-many alternation in auditory stimulus A. The signal indicated in blue represents the filtered output. The signal indicated in red represents the FFT-smoothed and scaled representation of the blue signal.

4.2 Movement feature extraction

The movement feature of interest is related with the amount of contraction/expansion of the upper body, which is here expressed by the changing distances between (1) the elbows relative to each other, and (2) the wrists relative to each

other. An increase in distance means an expansion of the used peripersonal space and vice versa.

The amount of contraction/expansion is calculated as the Euclidian distance between positions, according to equation 7, where d is distance, and ΔE is the difference in position between the left (E_1) and right elbow (E_2) in (x, y, z) .

$$d(t) = \sqrt{\Delta E_x(t)^2 + \Delta E_y(t)^2 + \Delta E_z(t)^2} \quad (7)$$

The signal is then scaled between 0 (i.e. minimum distance between the elbows) and 1 (i.e. maximum distance between the elbows) and smoothed with a Savitzky-Golay FIR smoothing filter of polynomial order 3 and with a frame size of 151. The same operations are executed to obtain the contraction/expansion index of the wrists.

4.3 Cross-correlation between movement feature and sensory feature

The comparison between (1) extracted movement features (distance between elbows, or wrists), and (2) auditory stimulus (regularly repeated alternations between tonic, tonic + third, and tonic + third + fifth) is based on a cross-correlation analysis, from which the highest correlation coefficient r is selected within a given time lag interval. A time lag interval is allowed in order to take into account the anticipation/retardation behaviour of the subjects. However, the value of the time lag is limited to a quarter of a period that characterizes an alternation between tonic, tonic + third, and tonic + third + fifth and back. This is done in order not to cancel anti-phase correlation patterns.

Every subject ($N=25$) performed the motor-attuning task on each of the four auditory stimuli. From the captured movement data of each performance, features were extracted and cross-correlated with the signal that describes the structural musical feature in each stimulus. In this way, for each of the auditory stimuli, 25 correlation coefficients were obtained relative to the performances of each subject. These data can be structured in a 25-by-4 data matrix wherein each column bundles the 25 correlation coefficients of each auditory stimulus. Once this data structure was obtained a statistical distribution was fit to the 25 correlation coefficients of each of the four columns. These four fitted distributions, represented by probability distribution functions (PDF), can be seen in figure 4 (above). The horizontal X-axis locates all the possible values of the correlation coefficient expressing the correlation between movement and musical feature while the vertical Y-axis describes how likely these values can occur (i.e. the probability) for each auditory stimuli. The same operations were executed for the movement feature that defines the distance between the two wrists (figure 4, below). Table 2 gives an overview of the means and medians of the eight different distributions.

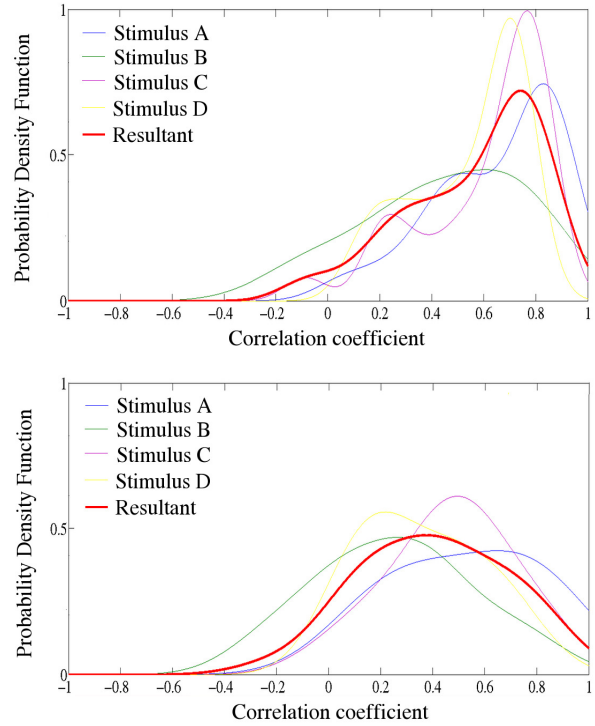


Fig. 4 Above: Scaled PDF expressing the probability distribution for each auditory stimulus of the correlation coefficients defining the similarity between movement feature (i.e. the distance between the elbows) and musical feature (the one-to-many alternation). Below: the same for the movement feature defining the distance between the two wrists.

Table 2 The means and medians of the 25 correlation coefficients r of each of the four data vectors that express the similarity between the varying distance of the elbows/wrists and the one-to-many alternation in the four musical stimuli)

Stimuli	A	B	C	D	All
N	25	25	25	25	100
Elbow					
r_{mean}	0.65	0.45	0.60	0.54	0.56
r_{median}	0.76	0.46	0.71	0.64	0.65
Wrist					
r_{mean}	0.53	0.25	0.47	0.36	0.40
r_{median}	0.52	0.24	0.48	0.28	0.41

Analysis of the correlations shows that the coordination between the movements of the wrists and the music is significantly smaller than the coordination of the elbows with the music. Analysis of variance shows a significant difference for stimuli B ($F(1, 48) = 5.66, p < 0.05$) and D ($F(1, 48) = 8.62, p < 0.01$), while there is a similar tendency for the two other stimuli, with p -values of 0.10 and 0.06 for A and C respectively. A significant difference between the four stimuli is found for both the wrists ($F(3, 96) = 5.48, p < 0.01$) and the elbows ($F(3, 96) = 2.87, p < 0.05$). Post-hoc tests show that in both cases the difference is caused by a poorer coordination in stimulus B.

These observations suggest, in general, that the elbows as well as the wrists (although in a lesser degree) attune with the auditory stimuli. This means that, when extra harmonic voices are added to a monophonic voice there is a general tendency to make broader gestures and vice versa. This indicates that a gestural trajectory is established as a response to the one-to-many alternation, expressed as a pattern of contraction/expansion of the upper limbs in peripersonal space. However, this seems not the case for every performance. From the lower ends of the different probability curves, it can be observed that some performances do not establish a clear sensorimotor coupling. Therefore, the results of the performances were related to the music and dance background of the subjects in order to see if this influenced the performance results. For each of the 25 subjects, a mean was calculated of the eight correlation coefficients defining how well the distances between both the elbows and both the wrists correlated with the one-to-many alternation in each of the four sound stimuli. The distribution of these 25 mean correlation coefficient in relation to the years of musical and dance education that the subjects experienced, revealed no linear relationship (see figure 5). The relation was measured by calculating the correlation coefficient between the variable specifying the correlation between movement and music and the two variables *years of music education* and *years of dance education* (respectively $r = 0.30$ and $r = 0.09$).

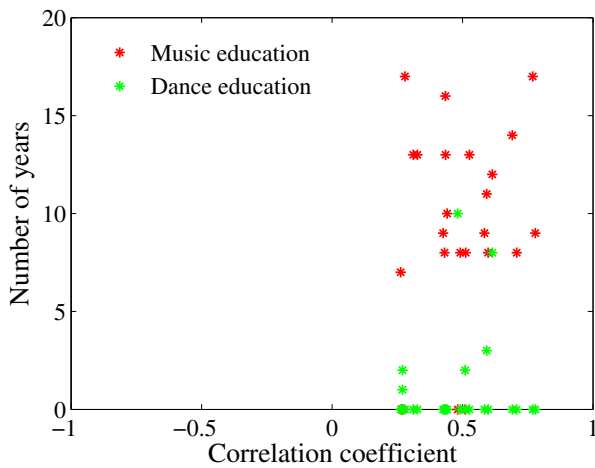


Fig. 5 Distribution of mean correlation coefficients in relation with artistic (i.e. music and dance) background of subjects.

An alternative analysis is presented in section 5 focussing on the subjective descriptions that subjects gave after each performance in order to find out whether these could be related with the performance results.

5 Quality of experience

The subjective descriptions of the participants experiences were given after each measurement, which gives a total of 100 ($=N25*4$) description moments, and a total of 204 descriptors. As these descriptions were given on a free and spontaneous basis, they were first categorized into four categories, namely, expressivity, richness, intensity and structure. These description types go together with a model for semantic description of music used by Lesaffre [29] in an experiment that focused on unveiling relationships between musical structure and musical expressiveness. The category expressivity (11%) relates to interest, affect and judgment and in our test it was expressed with adjectives such as irritancy, cheerfulness and difficulty. Richness (17%) consists of tone qualities such as fullness, wealth and spatiality. Intensity (9,3%) was expressed in terms of exuberance, intensity, force, and impressiveness. Structure (63%) had a focus on aspects of harmony (e.g. from unison to many-voiced, triadic) and movement (e.g. larger, wider, repetitive). The large amount of structure-related adjectives can be explained by the interview setup, which requires participants to give adjectives that express participants perception of evolution within the musical samples and their overall experience during the performance. Starting from this categorical approach a quality of experience value was assigned to each of the 100 description sets, using a score from 0 to 3. This was done twice by two musicologists who worked independently of each other. Their assessments resulted in two spreadsheets that were qualitatively compared with each other. In 93% of the cases, an agreement was found between both the assessments. In the other cases, disagreement occurred due to specific assessment of the terms *greater* and *louder*. After discussion and based on Lesaffres [29] model for semantic description of music, they were put in the *intensity* category. The obtained scores lead to four quality of experience groups in the following way: (1) the lowest score of 0 (5 descriptors), groups the description sets that provide no, minimal or vague sketches of the sensory stimuli; (2) the score of 1 (17 descriptors) has been given to the group that holds to one single descriptor category. This group also represents subjects who among other things said that they focused on the pulse (rhythm) of the sound stimuli, had negative experiences such as irritancy and oppression, or found the task difficult; (3) the group with a score of 2 (49 descriptors) corresponds with rather well described experiences, but the description simply accounts for one or two out of four description categories; (4) the group with the highest score of 3 (29 descriptors) consists of description sets that include multiple descriptors spread over the four categories, although there were no descriptions sets that covered all four description categories. The distribution of correlation coefficients (representing the probability of correlation between movement

and auditory stimulus) was then re-calculated for each of the four quality of experience groups. Once all the values of the correlation coefficients are obtained, a fitted probability distribution (PDF) is created for each group (see figure 6).

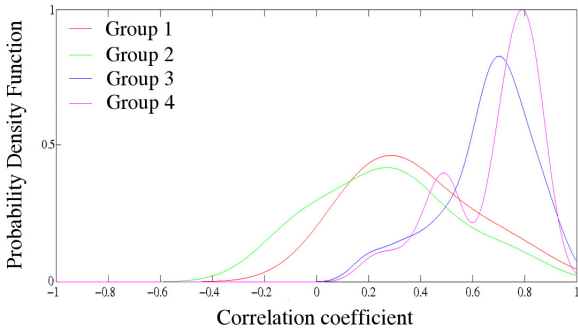


Fig. 6 Scaled probability distributions (PDF) that define the degree of sensorimotor coupling (expressed with correlation coefficient values that determine the correlation between the expansion/contraction pattern of the elbows and the one-to-many alternation in the music) as a function of the quality of experience. Four groups are defined based on different qualities of experience in terms of intensity, richness, expressivity and structure.

With a mean r of respectively 0.37 and 0.27, group 1 and 2 score less good than group 3 and 4 who have a mean r of respectively 0.65 and 0.67. These results suggest that when subjects are capable of giving a clear description of their experience in terms of expressive features like richness and intensity and structural features, they are more likely to give proof of a sensorimotor coupling (i.e. motor resonance behaviour).

The mapping trajectory that will be proposed further in this paper is based on the results of the performances of which the subjective descriptors were categorized in group 3 and 4 (see figure 6). In contrast with the performances that have a subjective descriptor in group 1 and 2 (i.e. 22 performances), performances of group 3 and 4 (i.e. 78 performances) gave evidence of a strong sensorimotor coupling accompanied with clear descriptions of subjective percepts in terms of coherent expressive and structural features. As a result, only the 78 performances classified in group 3 and 4 are relevant and meaningful for the development of the mapping trajectory.

Figure 7 shows how the remaining 78 performances can be distributed according to the degree of correlation between (1) the expansion/contraction pattern of the upper limbs into peripersonal space and (2) the one-to-many alternation in the pre-recorded auditory stimuli. This statistical distribution of the 78 correlation coefficients concerned with the amount of

expansion/contraction at the level of the elbows is characterized by a mean r of 0.66 (median $r = 0.71$). With a mean r of 0.47 (median $r = 0.45$), the correlation between the expansion/contraction pattern of the wrists in reference to the auditory stimuli is existing although less convincing. So, it seems that motor imitation occurs especially at the level of the elbows.

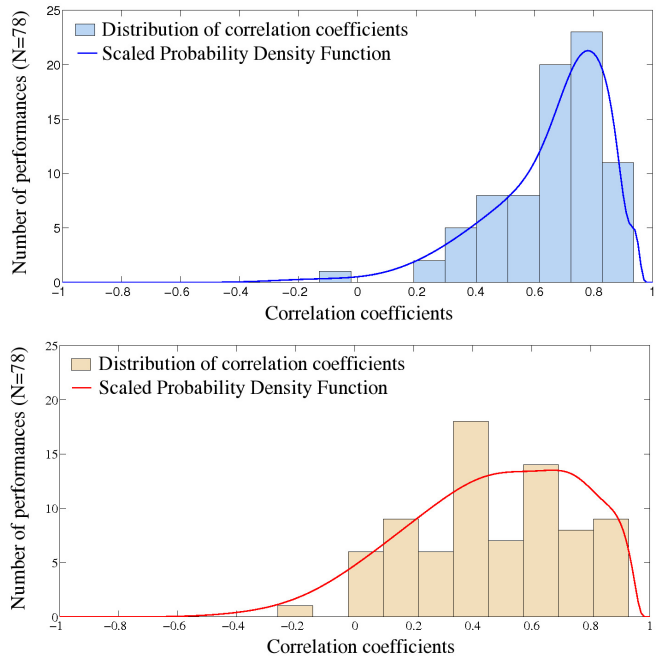


Fig. 7 Distribution and PDF of the correlation-coefficients ($N=78$) expressing the correlation between the varying distance between the elbows (above) and wrists (below) and the one-to-many alternation in the music.

6 Analysis of the directionality of the expansion in peripersonal space

The previous section indicated that an attuning is established between movement and sound defined by a contraction/expansion pattern of the upper limbs. However, this expansion can be the result of different movements of the upper limbs. So further analysis of position-related aspects needs to be conducted in order to specify how this spatio-kinetic movement trajectory is defined with respect to the 3D peripersonal space of the subjects. This analysis needs to incorporate the trajectory of the elbows and wrists along the different axes of the 3D coordinate system during each performance.

The trajectory along the axes is approached from the perspective of the subjects' peripersonal spaces. Therefore, the trajectory along the X-axis is scaled from 0 to 1 enabling an optimal comparison with the signal that specifies the dis-

tance between the elbows. The 0-value corresponds with the 0-value position on the horizontal X-axis, the 1-value corresponds with the maximum distance of the elbows and wrists to the body in both directions along the horizontal X-axis. The trajectory along the Z-axis is defined according to the same concept but in the horizontal forward/backward direction. The trajectory along the vertical Y-axis has a 0-value at the lowest possible position of the elbow and wrist and a 1-value at the highest position. It must be stated that a 1-value has a different length according it concerns the elbow or the wrist.

For each performance (N=78) correlations were made between (1) the trajectories of both the elbows and wrists along the three different axes and (2) the varying distance between the elbows. In this way, its possible to estimate the spatio-kinetic trajectory that determines the motor resonance behaviour of subjects in response to the one-to-many alternation in the auditory stimuli. After these correlations were executed, a matrix was created consisting of 12 columns of each 75 correlation coefficients. Departing from the statistical distribution of each column, means, medians and percentiles (25th and 5th) were calculated (table 3). Moreover, from each statistical distribution, there was a fitted probability distribution obtained represented by a PDF (figure 8). These statistical calculations give an indication of how the data values are spread over the interval from the smallest value to the largest value of the correlation coefficient. The results indicate that there is not so much difference between the left and right parts of the upper body. Because of that, in what follows the difference between the two body parts will be excluded and mean values will be used.

A first global observation of the means, medians, percentile values and PDF plot suggests that there are high r values concerning the trajectories of elbows and wrists along the vertical Y-axis. For the elbows, 95% of the r values (i.e. 74 out of the 78 values in total) go beyond 0.56. The probability density function indicates that the r value that is most likely to occur amounts to 0.93. For the wrists, the r value that seems most likely to occur amounts to 0.90. 95% of the r values are higher than 0.33 while 75% of the r values (i.e. 59 out of the 78 values in total) exceed 0.69. These observations suggest that the motor response behaviour determined by an expansion of the elbows is largely dependent on a displacement of the elbows and wrists in the upward direction. Further analysis is executed in order to investigate whether the upward movement of the elbows is due to a horizontal movement in the forward/backward direction (i.e. Z-axis) or in the sideward direction (i.e. X-axis). In order to realize this, it was investigated for all the performances whether an increase in distance between the elbows corresponded also to a similar, increasing displacement of the elbows and wrists along the X-axis. For the left elbow the con-

traction/expansion range in the horizontal, sideward direction is defined by x-coordinate values that fall in between 0 (i.e. maximum contraction) and 0.43 (i.e. maximum expansion). For the right elbow, the contraction/expansion range is defined by x-coordinates falling in between 0 and -0.43. For the left wrist, the x-values fall in between 0 and 0.7, for the right wrist, in between 0 and -0.7. In order to simplify the interpretation of the subsequent correlation analysis, the x-values for the right elbow and wrist were multiplied by -1 in order to relate increasing x-values to an expansion of the elbows and wrists. Moreover, the different ranges were normalized between 0 and 1. Four different correlation analyses were then applied between the two variables defining (1) the scaled x-values specifying the displacement of both the elbows and wrists and (2) the varying distance between both the elbows. For the elbows, 95% of the r values exceed 0.37, while 75% of the r values go beyond 0.62. The PDF fit of the distribution indicates that a value of 0.84 is most likely to occur. What concerns the wrists, the value that is most likely to occur is 0.43. These results suggest that the tendency towards spatial verticality is related with an expansion of the upper limbs in the outward direction along the X-axis. A similar correlation analysis was performed between the variables specifying the displacement along the Z-axis and the distance between the elbows. This analysis indicated an absence of correlation between the expansion of the elbows and the displacement of both the elbows and wrists along the Z-axis.

Table 3 Means, medians and percentile values of the 12 distributions of 78 correlation coefficients r that define the correlation between (1) the expansion of the elbows and (2) the displacement of the upper limbs along the three axes in peripersonal space. The six columns above contain the correlation of the expansion of the elbow with the trajectories along the X, Y and Z-axes of the left and right elbow. The six columns below these of the wrists.

N	78	78	78	78	78	78
Direction	X	X	Y	Y	Z	Z
Elbow	left	right	left	right	left	right
r_{mean}	0.78	0.68	0.90	0.85	0.09	0.40
r_{median}	0.78	0.73	0.91	0.92	0.11	0.39
$r_{prctl(25)}$	0.70	0.53	0.87	0.83	-0.09	0.18
$r_{prctl(5)}$	0.50	0.24	0.68	0.43	-0.40	-0.10
Wrist	left	right	left	right	left	right
r_{mean}	0.50	0.30	0.79	0.74	0.16	0.25
r_{median}	0.55	0.28	0.84	0.85	0.13	0.24
$r_{prctl(25)}$	0.32	0.14	0.71	0.66	-0.07	0.05
$r_{prctl(5)}$	-0.01	-0.13	0.43	0.22	-0.32	-0.14

7 Discussion

Similar with previous findings [34, 17, 18, 12], our study indicates that subjects have an intuitive tendency to associate

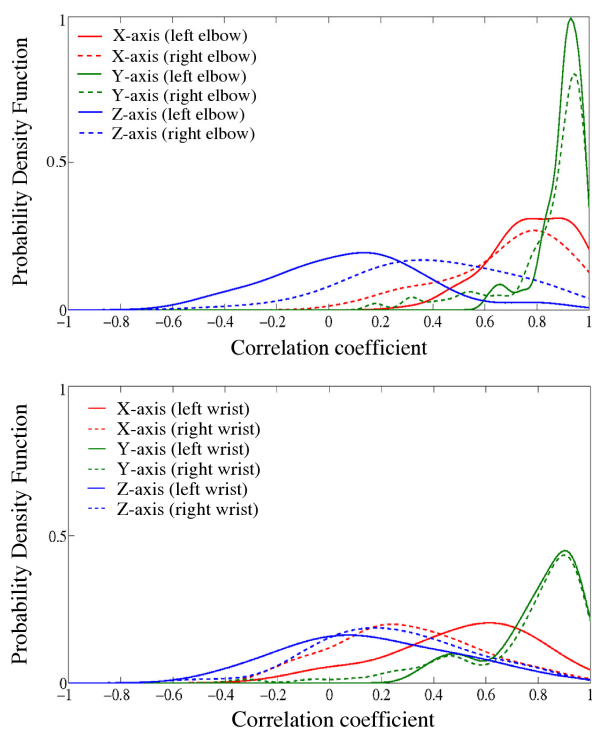


Fig. 8 Scaled fitted probability distributions (PDF) that define the correlation between (1) the expansion of the elbows and (2) the displacement of the elbows (above) and wrists (below) along the three axes in peripersonal space.

musical structures with physical space and bodily motion. The analysis of the extracted movement features shows that the corporeal resonance behaviour in response to the one-to-many alternation in the auditory stimuli is characterised by a contraction/expansion of the upper limbs in the subjects' peripersonal space. This seems to be the case especially what concerns the elbows. A partial explanation of this result could be that the wrists have rather a tendency to follow the regular pulse of the musical stimuli. A first brief, qualitative observation of the data seems to confirm this explanation although further research is needed.

Moreover, empirical evidence shows that the expansion of the upper limbs is characterized by an upward (Y-axis) and (although in a lesser degree) sideward (X-axis) directionality. Eitan and Granot [17] who investigated the relationship between musical structural descriptions and synesthetic/kinesthetic descriptions found that changes in pitch height corresponds to changes in movement characterised by a spatial vertical directionality as well. This could raise the question whether it is the addition of extra voices that stimulates expansion or whether it is the change in pitch. The experimental set-up anticipated this question. By adding voices in the octave below the root voice, there could be no question of a raising pitch in stimuli A and B. Nevertheless, there is a clear tendency to spatial verticality.

Furthermore, analysis suggested that there is no linear relationship between the artistic (i.e. music and dance) background of subjects and their motor-attuning performances on the pre-recorded stimuli. An alternative analysis however confirms the influence of personality and contextual factors, like nervousness and uneasiness, as important influences on the performance of subjects. These two observations feed the hypothesis that the observed corporeal resonance behaviour (in the case when a person is feeling comfortable) is really a spontaneous one, rooted in the action-oriented ontology of subjects independent of their artistic background. However, to confirm this hypothesis more adequate, additional research needs to be conducted with more and more divergent subjects preferably in more ecological valid situations.

The results of this study can be taken as ground for the establishment of a mapping trajectory that connects the analysed gestural trajectory of the upper limbs in peripersonal space defined by the contraction/expansion index to a module that controls the synthesis of extra voices on a monophonic auditory input (e.g. MuVoice module). In order to realize this, the computational model proposed in this paper must be adjusted to recognize the gestural trajectories of users, equipped with five inertial sensors and connected with the system. Once a pattern of contraction/expansion of the upper limbs is then recognized, the computational model must activate a sound-synthesizing module that adds voices on the monophonic input. We propose a mapping trajectory that (1) gradually adds a third according to the expansion of the elbows in the upward (Y-axis) and sideward (X-axis) direction and (2) gradually adds an extra fifth on the root and third according to the expansion of the wrists in the upward (Y-axis) and sideward (X-axis) direction. However, this proposed strategy must be validated in the future in ecological valuable environments and situations.

The innovative aspect of this mapping strategy is that it relies on an embodied, sensorimotor approach to music perception and expressiveness putting in evidence the action-oriented ontology of users [22, 32, 10, 39]. According to this approach, perception is build upon structural couplings between sensory patterns and motor activity. The qualitative research presented in this study (see section 5) confirms this these by showing that only the subjects that give evidence of a structural coupling between (1) the auditory pattern (i.e. one-to-many alternation) inherent in the auditory stimuli and (2) the corporeal resonance behaviour, could give a clear and coherent description of how they perceived the musical stimuli. In contrast, the subjective descriptions of the participants lacking the sensorimotor coupling were mostly vague and incoherent. Moreover, the subjective descriptions of the subjects that gave proof of a sensorimotor coupling were very much in accordance with each other. They described the passage from one voice to more, harmonic re-

lated and simultaneous sounding voices in terms of more full, more intense, more forceful, more exuberant, etc. As a result, a connection could be established between musical structural feature, movement feature and the perceptual experience in terms of richness, intensity and expressivity (table 4). This observed relationship is in line with previous research putting in evidence the relation between (1) the size and openness of bodily movement and the emotional intensity of the musical sound [14, 7] and (2) the contrast between an open and closed body position and the intent of communicators to persuade [31, 30].

Table 4 Relationship between musical structural feature, movement feature and perceptual experience in terms of richness, intensity and expressivity.

Structural musical cue	Movement cue	Perception
adding voices	expansion	full, intense, force
dissappearing voices	contraction	empty, delicate

By developing a mapping strategy that relies on sensorimotor integration and the coupling of action and perception, it seems possible to address descriptions on a pure subjective level (ideas, feelings, moods, etc.) by means of corporeal articulation and translate them further into sound. It must be noticed that the proposed sound process of one-to-many alternation is not related with one particular emotion or affect as such but rather with the intensity or quality of *an* emotion whatsoever [36, 8]. Nonetheless further research is required, the results presented in this paper suggest that an emotion, affect or dramatic content could be reinforced by the specific musical process of adding voices to an in origin monophonic voice. By integrating a musical synthesis module that is able to perform this kind of musical process (MuVoice) inside the algorithm that is able to model movement, a multimodal, digital interface is created that can be used by singers to enhance the natural capabilities of the voice to communicate emotions and feeling by means of the expressive qualities of the upper body extended into the virtual musical domain. In this way, it also provides a tool that enhances music-driven psychosocial interaction in an artistic and social context. It facilitates users, multimedia artists and dancers to control and manipulate a music performance by means of expressive gestures. For the performer it means that the performed actions are attuned with the intended actions communicated by the musical structural cues in the constitution of unambiguous expressive percepts. For the outside world, it means that visual and auditory sensory input received from stage (visual, auditory,...) are perceived as attuned intended actions that can be corporeally imitated and related to the action-oriented ontology creating expressive content.

However the advantages of the proposed mapping strategy, further research needs to be conducted in order to in-

tegrate the proposed methodology into systems that support one-to-many, many-to-one or many-to-many mapping strategies [35, 21]. This will be accomplished in the future by conducting additional experiments investigating the relation between expressive gesture and sound.[33]

Acknowledgements This work is funded by the EmcoMetecca project Ghent University.

We want to thank Pieter Coussement for his contribution to the project with the development of a Jitter-generated stick-figure representation. We also want to thank Mark T. Marshall and Marcelo Wanderley at the Input Devices and Music Interaction Laboratory at McGill University (www.idmil.org) for their software that accesses the Xsens sensors and converts the received data to the OSC protocol.

References

1. N. Bernardini. <http://www.cost287.org/>.
2. F. Bevilacqua, J. Ridenour, and D.J. Cuccia. 3D motion capture data: motion analysis and mapping to music. In *Proceedings of the Workshop/Symposium on Sensing and Input for Media-centric Systems*, 2002.
3. N. Bianchi-Berthouze, P. Cairns, A. Cox, C. Jennett, and W.W. Kim. On posture as a modality for expressing and recognizing emotions. In *Emotion and HCI workshop at BCS HCI London*, 2006.
4. C. Cadoz and M.M. Wanderley. Gesture-music. *Trends in Gestural Control of Music*, pages 71–93, 2000.
5. A. Camurri, G. De Poli, A. Friberg, M. Leman, and G. Volpe. The MEGA project: analysis and synthesis of multisensory expressive gesture in performing art applications. *Journal of New Music Research*, 34(1):5–21, 2005.
6. A. Camurri, G. De Poli, M. Leman, and G. Volpe. A multi-layered conceptual framework for expressive gesture applications. In *Proc. Intl MOSART Workshop, Barcelona*, 2001.
7. A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers, and G. Volpe. Multimodal analysis of expressive gesture in music and dance performances. *Lecture notes in computer science*, pages 20–39, 2004.
8. A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers, and G. Volpe. Multimodal analysis of expressive gesture in music and dance performances. *Lecture notes in computer science*, pages 20–39, 2004.
9. A. Camurri, G. Volpe, G. De Poli, and M. Leman. Communicating expressiveness and affect in multimodal interactive systems. *IEEE Multimedia*, 12(1):43–53, 2005.
10. G. Colombetti and E. Thompson. The feeling body: Toward an enactive approach to emotion. *Body in mind, mind in body: Developmental perspectives on embodiment and consciousness*. Hillsdale, NJ: Lawrence Erlbaum, 2007.
11. M. Coulson. Expressing emotion through body movement. *Animating Expressive Characters for Social Interaction*, page 71, 2008.
12. P. Craenen. Music from Some (no) where, Here and There: Reflections over the Space of Sounding Compositions. *TIJDSCHRIFT VOOR MUZIEKTHEORIE*, 12(1):122, 2007.
13. S. Dahl. *On the beat: Human movement and timing in the production and perception of music*. PhD thesis, KTH School of Computer Science and Communications, SE-100 44 Stockholm, Sweden, 2005.
14. J.W. Davidson. What type of information is conveyed in the body movements of solo musician performers. *Journal of Human Movement Studies*, 6:279–301, 1994.

15. P.R. De Silva and N. Bianchi-Berthouze. Modeling human affective postures: an information theoretic characterization of posture features. *Computer Animation and Virtual Worlds*, 15, 2004.
16. PR De Silva, M. Osano, A. Marasinghe, and AP Madurapperuma. Towards recognizing emotion with affective dimensions through body gestures. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 269–274, 2006.
17. Z. Eitan and R.Y. Granot. Musical parameters and images of motion. In *Proceedings of the Conference on Interdisciplinary Musicology (CIM04), Graz/Austria*, pages 15–18, 2004.
18. Z. Eitan and R.Y. Granot. How music moves. *Music perception*, 23(3):221–248, 2006.
19. A. Farne, M.L. Dematte, and E. Ladavas. Neuropsychological evidence of modular organization of the near peripersonal space. *Neurology*, 65(11):1754–1758, 2005.
20. D. Fenza, L. Mion, S. Canazza, and A. Roda. Physical movement and musical gestures: a multilevel mapping strategy. *Proceedings of Sound and Music Computing'05*, 2005.
21. A. Hunt, M. Wanderley, and R. Kirk. Towards a model for instrumental mapping in expert musical interaction. In *International Computer Music Conference*, pages 209–212, 2000.
22. SL Hurley. *Consciousness in action*. Harvard Univ Pr, 2002.
23. AR Jensenius. *Action-Sound: Developing Methods and Tools to Study Music-Related Body Movement*. PhD thesis, Ph. D. thesis, Department of Musicology, University of Oslo, 2007.
24. A. Kleinsmith, T. Fushimi, and N. Bianchi-Berthouze. An incremental and interactive affective posture recognition system. In *International Workshop on Adapting the Interaction Style to Affective Factors, Edinburgh, UK*, 2005.
25. G. Kurtenbach and E.A. Hulteen. Gestures in Human-Computer Communication. *The Art of Human-Computer Interface Design*, pages 309–317, 1990.
26. R. Laban and F.C. Lawrence. *Effort*. Macdonald & Evans London, 1947.
27. M. Leman. *Embodied music cognition and mediation technology*. Mit Press, 2007.
28. M. Leman and A. Camurri. Understanding musical expressiveness using interactive multimedia platforms. *Musicae Scientiae*, 10(I):209, 2006.
29. M. Lesaffre, L.D. Voogdt, M. Leman, B.D. Baets, H.D. Meyer, and J.P. Martens. 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(5), 2008.
30. H. McGinley, R. LeFevre, and P. McGinley. The influence of a communicator's body position on opinion change in others. *Journal of Personality and Social Psychology*, 31(4):686–690, 1975.
31. A. Mehrabian and J.T. Friar. Encoding of Attitude by a Seated Communicator via Posture and Position Cues. *J Consult Clin Psychol*, 1969.
32. A. Noë. *Action in perception*. MIT Press, 2004.
33. F. Offli, Y. Demir, Y. Yemez, E. Erzin, A.M. Tekalp, K. Balci, İ. Kızıoğlu, L. Akarun, C. Canton-Ferrer, J. Tilmanne, et al. An audio-driven dancing avatar. *Journal on Multimodal User Interfaces*, 2(2):93–103, 2008.
34. B.H. Repp. Music as motion: A synopsis of Alexander Truslit's (1938) *Gestaltung und Bewegung in der Musik*. *Psychology of Music*, 21(1):48, 1993.
35. J.B. Rován, M.M. Wanderley, S. Dubnov, and P. Depalle. Instrumental gestural mapping strategies as expressivity determinants in computer music performance. In *Proceedings of Kansei-The Technology of Emotion Workshop*, pages 3–4, 1997.
36. KR Scherer. Why music does not produce basic emotions: pleading for a new approach to measuring the emotional effects of music. In *Proc. Stockholm Music Acoustics Conference SMAC-03*, pages 25–28, 2003.
37. K. Schindler, L. Van Gool, and B. de Gelder. Recognizing emotions expressed by body pose: A biologically inspired neural model. *Neural Networks*, 21(9):1238–1246, 2008.
38. L. Tarabella and G. Bertini. About the Role of Mapping in gesture-controlled live computer music. *Lecture notes in computer science*, pages 217–224, 2004.
39. D. Taraborelli and M. Mossio. On the relation between the enactive and the sensorimotor approach to perception. *Consciousness and Cognition*, 17(4):1343–1344, 2008.
40. R. von Laban and FC Lawrence. *Effort*. Macdonald & Evans, 1967.
41. T. Winkler. Making motion musical: Gesture mapping strategies for interactive computer music. In *Proceedings of the 1995 International Computer Music Conference*, pages 261–264, 1995.