

## THE COUPLING OF ACTION AND PERCEPTION IN MUSICAL MEANING FORMATION

---

PIETER-JAN MAES, EDITH VAN DYCK,  
MICHELINE LESAFFRE, MARC LEMAN  
*Ghent University, Ghent, Belgium*

PIETER M. KROONENBERG  
*Leiden University, Leiden, The Netherlands*

**THE EMBODIED PERSPECTIVE ON MUSIC COGNITION** has stressed the central role of the body and body movements in musical meaning formation processes. In the present study, we investigate by means of a behavioral experiment how free body movements in response to music (i.e., action) can be linked to specific linguistic, metaphorical descriptions people use to describe the expressive qualities they perceive in the music (i.e., perception). We introduce a dimensional model based on the Effort/Shape theory of Laban in order to target musical expressivity from an embodied perspective. Also, we investigate whether a coupling between action and perception is dependent on the musical background of the participants (i.e., trained versus untrained). The results show that the physical appearance of the free body movements that participants perform in response to music are reliably linked to the linguistic descriptions of musical expressiveness in terms of the underlying quality. Moreover, this result is found to be independent of the participants' musical background.

*Received: July 25, 2012, accepted January 24, 2013.*

**Key words:** embodied music cognition, expressive gestures, action-perception coupling, musical metaphor, musical expressiveness

**W**HEN PEOPLE SPEAK ABOUT WESTERN ART music, metaphors are ubiquitous (Cox, 2011; Guck, 1994). Embodied, sensory-motor experiences are considered to play an important role in the use and understanding of metaphors (Crawford, 2009; Gibbs, Costa Lima, & Francozo, 2004). Metaphorical constructs (e.g., "life is a journey") tend to elucidate abstract ideas (e.g., "life") by relating them to concrete, sensory-motor experiences (e.g., "journey"). In the context of music, research has

indicated that a substantial part of musical metaphors relate to people's experience of physical motion and space (Cox, 2011; Johnson & Larson, 2003; Larson, 2012). Moreover, in the general paradigm of embodied music cognition, theories have been developed explaining how people perceive and understand music, at least partly, in terms of the body movements that music induces. Examples are the motor model of musical expressiveness (Leman, 2007), the motor-mimetic theory (Godøy, 2003), and the mimetic hypothesis (Cox, 2011). In the present study, we elaborate on the idea that the embodied experience of music is an important factor in people's use and understanding of musical metaphors.

People often respond to music by making body movements that reflect aspects of its melodic and rhythmic contours, or of the sound-producing gestures from which it originated (Godøy, 2010). This may instigate a sense of imagined participation with the production of the sound, an idea that is addressed in a broad range of musicological studies under different names, such as *imagined activity* (Maus, 1988), *kinaesthetic empathy* (Mead, 1999), *imaginary agency* (Levinson, 2006), *simulated control* (Leman, 2007), and *active perception* (Krueger, 2009). What all of these accounts have in common is their reference to a direct, sensory-motor engagement with music; to how music literally "moves" people, and to how they feel they are immersed in, and resonate with, the physical sound energy. In the present study, we ask how this form of embodied sensory-motor engagement with music might be related to the use of musical metaphors. We present an experiment designed to investigate the extent to which physical and expressive properties of music-induced body movements (i.e., action) are reflected in the linguistic metaphors that people use to describe the expressive qualities they recognize in the music (i.e., perception). Here, perception refers to the identification, organization, and interpretation of what is received through the senses (here, the auditory sense), and involves a combination of "bottom-up" sensory processing, and "top-down" processes (expectation, knowledge, etc.). Accordingly, perception itself pertains to a level of understanding that could be conceptualized as linguistic metaphors.

TABLE 1. Expressive Model Based on Laban's Effort/Shape Theory.

EFFORT CATEGORIES								SHAPE CATEGORIES			
WEIGHT		TIME		FLOW		SPACE		Spreading-Enclosing		Rising-Descending	
Strong	Light	Sudden	Sustained	Bound	Free	Direct	Indirect				
LINGUISTIC LABELS											
heavy	light	fast	slow	rigid	fluent	compact	airy	big	small	high	low
vigorous	frail	nervous	tranquil	anxious	secure	dense	diffuse	broad	narrow	vertical	horizontal
rough	delicate	energetic	soothing	worried	carefree	regular	chaotic	thick	thin	dominant	humble
hard	soft	active	passive	serious	playful	balanced	unbalanced	exalting	serene	bright	muted
MOVEMENT FEATURES											
Acceleration		Impulsiveness		Smoothness		Directness Index (DI)		Size		Height	
high	low	high	low	low	high	straight	indirect	big	small	high	low
		Jerk									
		high	low								

Note: Linguistic labels and movement characteristics highlighted in grey are the ones that participants associated with the heroic music, the others with the lyric music.

Our analytical framework for mapping low-level physical movement properties onto high-level expressive linguistic concepts will be the Effort/Shape model that originated in the Laban Movement Analysis (LMA) method (Laban, 1947; Laban & Ullmann, 1966/2011). This is an integrated conceptual system that relates physical movements to expressive qualities such as emotion, intentionality, and so on. By adapting the model, we are able to incorporate linguistic metaphors to our analysis. The Effort/Shape model has been applied in earlier studies of expressive movement responses to music (Camurri, Mazzarino, Ricchetti, Timmers, & Volpe, 2004; Maes, Leman, Lesaffre, Demey, & Moelants, 2010), although no attempts have been made to use it to clarify the relationship between free movement responses to music and the use of linguistic musical metaphors. As the name implies, the model consists of two main components: *Effort*, which refers to the subtle, energetic qualities of an expressive movement; and *Shape*, which relates to changes in body shape. Both components are subdivided into different categories. There are four categories within the Effort component, and in turn these are structured around a conceptual opposition: 1) Effort-weight relates to the qualities *strong-light*; 2) Effort-time relates to *sudden-sustained*; 3) Effort-flow is structured around *bound-free*; and 4) Effort-space is classified as *direct-indirect*. Similarly, the Shape component is divided into three categories structured around the polarities *spreading-enclosing*, *rising-descending*, and *advancing-retreating*. This last opposition is not taken into account in the

present study as it is not relevant to hand gestures (see below). The crux of our study is to incorporate relevant movement features and linguistic metaphors to this model, reflecting the categories listed above and the oppositions around which they are structured (see Table 1). This enables us to study movements in response to, and the conceptualization of expressiveness within music, and to see how those two components correlate.

Our adapted version of the model can be described as follows: Effort-weight is related to the movement feature *acceleration*, and the bipolar indicators *heavy-light*, *vigorous-frail*, *rough-delicate*, and *hard-soft*. Effort-time is related to the movement feature *impulsiveness*, and the bipolar indicators *fast-slow*, *nervous-tranquil*, *energetic-soothing*, and *active-passive*. Effort-flow is related to the movement features *jerk* and *smoothness error* (i.e., measured as the deviation from the smoothest trajectory in m/sample), and the bipolar indicators *rigid-fluent*, *anxious-secure*, *worried-carefree*, and *serious-playful*. Jerk and smoothness error are related in the sense that the smoothness error is quantified as a function of jerk (cf. minimum jerk model; Todorov & Jordan, 1998). As the derivative of movement acceleration, jerk represents the rate of change in movement acceleration. Hence, “smoothness” is equivalent to small rates of change in movement acceleration (i.e., low jerk). Effort-space is related to the movement feature *directness index*, and the bipolar indicators *compact-airy*, *dense-diffuse*, *regular-chaotic*, and *balanced-unbalanced*. The directness index is calculated for individual segments of the motion

trajectories of the dominant hand. This measure expresses the ratio between the length of the straight trajectory connecting the beginning and end point of a motion segment and the distance of the actual covered trajectory between these two points. Shape-spreading/enclosing is related to the movement feature *size*, and the bipolar indicators *big-small*, *broad-narrow*, *thick-thin*, and *exalting-serene*. Shape-rising/descending is related to the movement feature *height*, and the bipolar indicators *high-low*, *vertical-horizontal*, *dominant-humble*, and *bright-muted*.

The experiment was conducted in two parts. In one part - the motor-attuning part - people were asked to perform free body movements in response to music, and their responses were recorded with an optical motion tracking system. In the other part - the self-report part - people were asked to associate 24 pairs of bipolar adjectives with the music while simply listening to the music. Rather than use various, shorter extracts of music, we opted for a single, continuous piece in which contrast is an inherent part of the composition. Dynamic contrasts are a typical feature of Late-Romantic music, and the particular music we chose was the beginning of the first movement of Brahms' *First Piano Concerto*. The reasons for this choice are threefold. First, listening to a continuous musical piece is assumed to create a more ecologically valid listening situation in which participants can immerse themselves in the music. Second, it allows us to investigate whether the expressive extremes of the material within a single piece are reflected in body movement responses and musical metaphors. Third, an emphasis on musical contrast corresponds to the polarities around which the Effort/Shape model is structured. We extracted acoustic properties from the audio signal; an energy property (i.e., *amplitude*) and spectrum properties (i.e., *irregularity*, *spectral flatness*, *spectral sharpness*, and *spectral variance*). Based on the contrasts we found in the musical material, we defined two musical styles: *heroic* and *lyric*. For each of these styles we identified three exemplary passages within the piece, and we asked participants to rate each of those six passages according to our set of 24 bipolar adjective scales. Our movement analysis took into account only those movements that were performed during those six specific passages, although participants were asked to perform free body movements continuously throughout the musical piece.

Three main research questions are asked in the present study. First, whether or not contrasts in music are reflected in people's movement responses to it, and in their ratings of its character across the bipolar adjective

scales. Second, whether there are similarities among the participants' responses, and whether these relate to the musical background of the participants (i.e., trained versus untrained). In addressing this question, we hypothesized that differences in movement responses and verbal descriptions between the two musical styles would be more pronounced in people with a musical background, as perception is influenced by learned, "top-down" processes involving expectation and knowledge. Third, using the Effort/Shape model, we wanted to investigate the correlation between people's movement responses and the musical metaphors they attribute to the music.

## Method

### PARTICIPANTS

We invited 36 students enrolled in a broad range of academic disciplines at Ghent University to participate in the experiment. They received no compensation for participating in the study. None of the participants received formal dance training in the past. Two distinct groups of participants were recruited systematically, on the basis of their musical background. The musically trained group comprised 18 participants (10 male, 8 female) with a mean age of 23.8 years ( $SD = 3.7$ ) and who played a musical instrument and had a mean number of 9.7 years ( $SD = 5.4$ ) of music training. The musically untrained group comprised 18 participants (10 male, 8 female) with a mean age of 24.6 ( $SD = 4.8$ ) and who did not have either music education of any sort (except for the obligatory courses in primary and secondary school) or play a musical instrument. In addition, we asked after the experiment how familiar participants were with the type of music used in the experiment. This question was rated on a five-point Likert scale, with 1 as *not at all familiar*, 3 as *somewhat familiar*, and 5 as *extremely familiar*. The mean rating of the participants in the musically trained group was 4.17 ( $SD = 0.38$ ), whereas the mean musical familiarity rating for the musically untrained group was 1.67 ( $SD = 0.48$ ).

We also tested the level of introversion/extraversion of each participant as we considered that this might be an influential factor. For this measure, we used the Dutch translation of the Big Five Inventory test (Denissen, Geenen, Van Aken, Gosling, & Potter, 2008) providing a score from 1 to 5, with 1 equalling *very introvert* and 5 equalling *very extravert*. A statistical *t*-test indicated that there was no significant difference between the two groups: for the musically trained group, ( $M = 3.49$ ,  $SD = 0.62$ ), and for the musically untrained group ( $M = 3.51$ ,  $SD = 0.67$ ),  $t(34) = -.10$ ,  $p > .05$  ( $p = .92$ ). So, both groups were highly homogeneous in this respect.

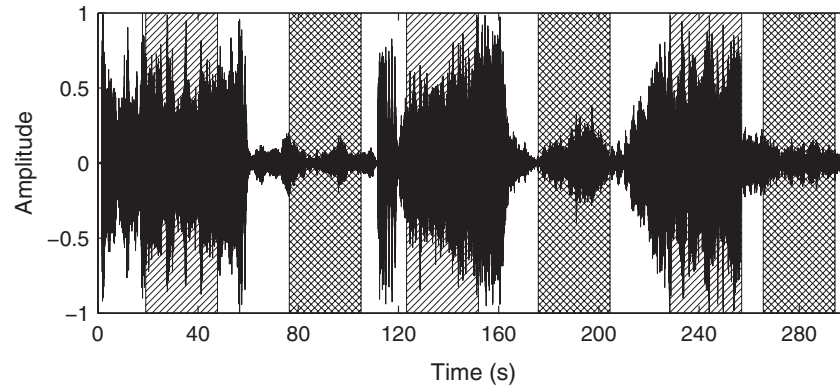


FIGURE 1. Waveform of the musical stimulus used in the experiment (i.e., the first five minutes of the first movement of Brahms' *First Piano Concerto*). The hatched regions (each 30 s long) indicate the three heroic fragments, the cross-hatched regions (each 30 s long) the three lyric fragments.

#### MATERIALS

*Musical stimulus.* The musical stimulus that was used in the experiment is Johannes Brahms' *First Piano Concerto*, Opus 15 in D minor from 1858. The recording we used was made by Krystian Zimmerman, accompanied by the Berlin Philharmonic Orchestra conducted by Simon Rattle. For the experiment, we used the first 6 min 10 s of the *Maestoso* movement (see Figure 1). In this excerpt three heroic passages alternate with three lyric passages (see Figure 1). The first of the lyric passages is relatively long in comparison with the other two, so we deleted a portion of that passage (1 min 56 s - 2 min 46 s of the recording), which was not evident for those unfamiliar with the piece. Thus, the eventual stimulus had a duration of 5 min 20 s. Fragments of equal duration were required for the movement analysis and self-report analysis, so we defined six fragments with a duration of 30 s each - three heroic and three lyric, as explained above. Contrary to what these labels might suggest, these two categories were not defined in terms of their expressive content, but purely in terms of the physical acoustic properties of the audio signal. We used the VAMP libXtract plugin (Bullock, 2007, 2008) hosted in Sonic Visualizer (Cannam, Landone, & Sandler, 2010) to extract the acoustic properties (see above). Precise details of the specific extraction methods can be found in Bullock (2008). For each acoustic property we obtained a continuous time series and then, for each one, the two distributions of samples (corresponding to the heroic and lyric fragments) were statistically analyzed and compared by means of a nonparametric Mann-Whitney *U* test. The results show that the levels of *amplitude*, *irregularity*, *spectral sharpness*, and *spectral variance* were significantly higher in the

heroic fragments compared to lyric fragments, and that the levels of *spectral flatness* were significantly lower in the heroic fragments compared to the lyric fragments ( $p < .001$ ).

*Questionnaire (self-report part).* The questionnaire was based on the semantic differential method (Osgood, Suci, & Tannenbaum, 1957), which can be used to assess people's attitudes towards all sorts of concepts, objects, events, etc. In particular, the method has proved useful in rating people's emotional experience of musical fragments (e.g., Fujihara & Tagashira, 1984; Murakami & Kroonenberg, 2003; Nielzén & Cesarec, 1981; O'Briant & Wilbanks, 1978; Senju & Ohgushi, 1987; Swanwick, 1973; White & Butler, 1968). The method is based on a set of scales consisting of bipolar pairs of adjectives, placed at each end of a continuous rating scale (usually a seven-point scale, although five-point and nine-point scales are also used). Respondents are asked to tick one of the seven points in between each bipolar adjective pair, indicating the rate of association (cf. intensity) of a particular concept, object, event, etc. with one of the adjectives (cf. direction). In our study, we used a 24-item, nine-point, semantic differential list to rate the six musical fragments, as we were interested in small nuances between different respondents. Each category in the Effort/Shape model itself is constituted by four bipolar adjective indicators (Table 1), and we chose our adjectives from the Laban Movement Analysis (LMA) effort bank (<http://www.lmaeffortbank.com>) as well as the Hevner adjective circle. The latter presents adjectives that are supposed to reflect the principal feelings that can be evoked in humans by listening to music (Hevner, 1936). Respondents were asked to rate the six musical

fragments while they listened to the corresponding music. The questionnaire was administered with Qualtrics ([www.qualtrics.com](http://www.qualtrics.com)) running online on a MacBook Pro.

#### PROCEDURE

Participants took part in the experiment individually, and beforehand, they were each given an explanation of the procedure and the tasks they had to carry out. Necessarily, they were equipped with the motion capture jacket and followed calibration procedure. The order in which the two parts of the experiment (motor-attuning part, self-report part) were executed was counterbalanced to eliminate order effects. Both parts of the study were carried out in the motion capture space, an octagonal area of 6.5 m diameter, enclosed by black curtains in order to separate the participants from the experimenters. Prior to the motor-attuning part of the experiment, the participants received the following instructions: “Translate your experience of the music into free full-body movement. Try to become absorbed in the music that is presented and express your feelings into body movement. There is no good or wrong way of doing it. Just perform what comes up in you.” They were allowed to use a space indicated by a round carpet, with a diameter of 4 m. We made the room completely dark, as the pilot study had indicated that this helped participants to feel more comfortable and less constrained in executing their task. The music was played through a stereo setup comprising two Behringer B2031A Truth Active Studio Monitors at a predefined, comfortable volume. In the self-report part of the experiment, participants filled in the questionnaire at a table adjacent to the carpet. They were asked to fill in the questionnaire while listening to the music. Again, the room was darkened and the same playback system was used, just as in the motor-attuning part of the experiment. Participants were instructed to tick one of the nine points between each bipolar adjective pair, indicating the level of association they perceived between the descriptors and each of the musical fragments. The order of the musical fragments was counterbalanced, to avoid order effects. From beginning to end, each experiment lasted for about one hour, at the end of which people were asked to fill in an exit questionnaire about their experience of the experiment. Finally, participants were offered refreshments while the experimenters explained the purpose of the experiment.

#### MOVEMENT DATA ACQUISITION (MOTOR-ATTUNING PART)

In our study, we focused on hand gestures. This choice was motivated by previous research showing that the

hands play a privileged role in music-induced gestures (Godøy, 2010). Three-dimensional movements of both hands were captured with an OPTITRACK infrared optical system consisting of 12 synchronized cameras and recorded with the ARENA motion capture software (<http://www.naturalpoint.com>). Participants were asked to wear a special jacket on which markers could be attached with Velcro. A default human upper body skeleton model provided in the ARENA software was constructed from 19 infrared reflecting markers that were attached to the jacket in a predefined manner: four markers for the hip and three markers for the chest, upper arms, and hands. Following the testing phase, data from the performances of all participants were exported into BioVision Hierarchy (BVH) files. Using the MATLAB motion capture toolbox (<http://www.cs.man.ac.uk/neill/mocap>) complemented with our own algorithms, we calculated the three-dimensional position and displacement of both hands in reference to the body-center (the center of the hips) independent of how the participant is positioned or orientated relative to the motion capture space. The selection and calculation of the individual movement features (acceleration, impulsiveness, smoothness error, jerk, directness index, size, height) was based on previous studies applying the Effort/Shape model in the study of human movement behavior (Camurri et al., 2004; Petersen, 2008; Van Dyck, Maes, Hargreaves, Lesaffre, & Leman, 2013). A full description of the different features and the procedures used to calculate them is presented in Van Dyck et al. (2013). Accordingly, for each subject ( $N = 36$ ) we obtained a single-value measure per musical style for each of the seven movement features. With these data, we drew up seven  $36 \times 2$  (i.e., participants  $\times$  styles) matrices for further statistical analysis.

## Results

#### MOTOR-ATTUNING TASK: MOVEMENT ANALYSIS

The seven  $36 \times 2$  matrices representing the general measures of all movement features were integrated in a  $2 \times 2 \times 2$  mixed-design ANOVA, with musical background (musically trained/untrained), and task order (first motor-attuning/self-report) as between-subjects factors and musical style (heroic/lyric) as a within-subject factor. The dependent variable, movement response, is in fact constituted by multiple dependent measures related to the different movement features under consideration (i.e., acceleration, impulsiveness, jerk, smoothness error, directness index, size, height). The main question concerns the effect of these dependent variables on changes within the two groups,

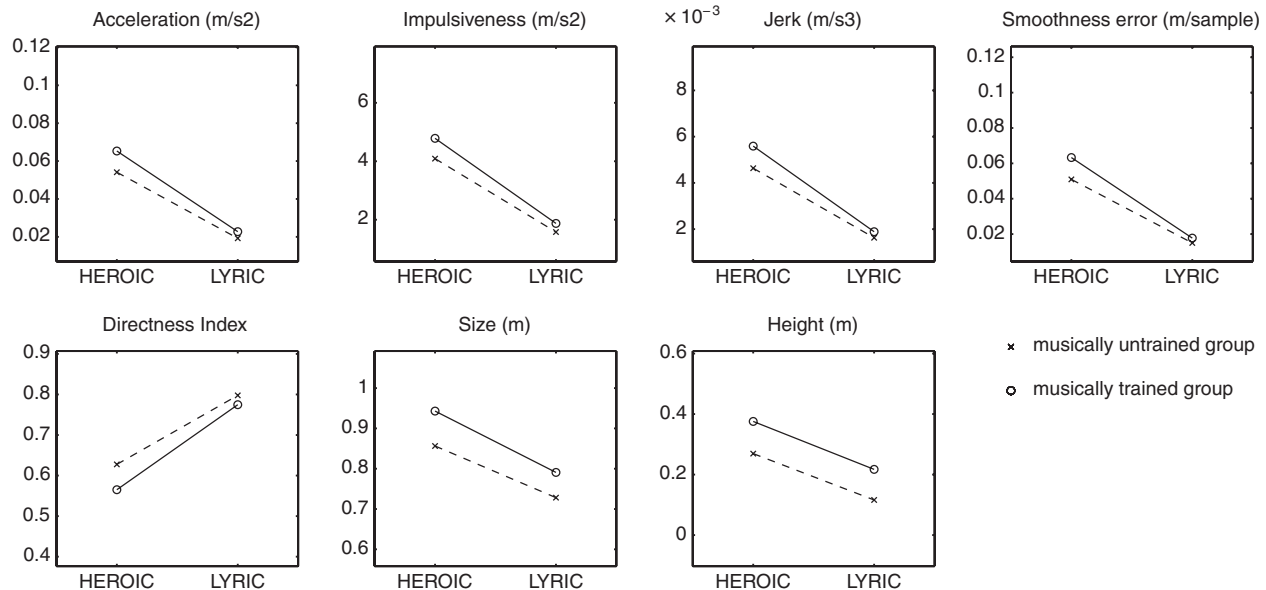


FIGURE 2. Representation of the means over all subjects ( $N = 36$ ), per group (musically trained/untrained), per condition (heroic/lyric) for all seven movement features.

rather than intercorrelations between the outcome variables. Therefore, it is appropriate to apply multiple univariate analyses on each of the dependent variables (ANOVA) instead of using a single multivariate analysis (MANOVA) (cf. Huberty & Morris, 1989; Park & Schutz, 2006).

Kolmogorov-Smirnov and Levene's tests showed that the assumptions of normality and homogeneity of variance could be accepted for all features ( $p > .05$ ). All effects displayed in Figure 2 are reported as significant at  $p < .001$ . There was a significant main effect of musical style for acceleration,  $F(1, 34) = 163.42$ ,  $r = .91$ , impulsiveness,  $F(1, 34) = 187.96$ ,  $r = .92$ , jerk,  $F(1, 34) = 171.36$ ,  $r = .91$ , smoothness error,  $F(1, 34) = 124.50$ ,  $r = .89$ , directness index,  $F(1, 34) = 125.38$ ,  $r = .89$ , size,  $F(1, 34) = 124.84$ ,  $r = .89$ , and height,  $F(1, 34) = 75.48$ ,  $r = .83$ . These results revealed that, in the lyric condition ( $M = 0.79$ ,  $SD = 0.06$ ), the level of the directness index was higher than in the heroic condition ( $M = 0.60$ ,  $SD = 0.11$ ). In addition, in the heroic condition, levels of acceleration ( $M = 0.06$ ,  $SD = 0.02$ ), impulsiveness ( $M = 4.44$ ,  $SD = 1.51$ ), jerk ( $M = 0.005$ ,  $SD = 0.002$ ), smoothness error ( $M = 0.06$ ,  $SD = 0.03$ ), size ( $M = 0.90$ ,  $SD = 0.12$ ), and height ( $M = 0.32$ ,  $SD = 0.15$ ) were higher compared to the lyric condition (acceleration:  $M = 0.02$ ,  $SD = 0.01$ ; impulsiveness:  $M = 1.73$ ,  $SD = 0.74$ ; jerk:  $M = 0.002$ ,  $SD = 0.001$ ; smoothness error:  $M = 0.02$ ,  $SD = 0.01$ ; size:  $M = 0.76$ ,  $SD = 0.11$ ; height:  $M = 0.17$ ,  $SD = 0.14$ ).

Regarding musical background, significant effects were found for size,  $F(1, 34) = 4.60$ ,  $p < .05$ ,  $r = .34$ , and height,  $F(1, 34) = 7.60$ ,  $p < .01$ ,  $r = .43$ : results for the musically trained group (size:  $M = 0.87$ ,  $SD = 0.12$ ; height:  $M = 0.30$ ,  $SD = 0.14$ ) were higher than those for the musically untrained group (size:  $M = 0.79$ ,  $SD = 0.14$ ; height:  $M = 0.19$ ,  $SD = 0.17$ ). However, no such effect was found for acceleration,  $F(1, 34) = 1.77$ ,  $r = .22$ , impulsiveness,  $F(1, 34) = 1.98$ ,  $r = .23$ , jerk,  $F(1, 34) = 1.70$ ,  $r = .22$ , smoothness error,  $F(1, 34) = 1.95$ ,  $r = .23$ , or directness index,  $F(1, 34) = 3.41$ ,  $r = .30$ , indicating that levels for these features were to all intents and purposes the same in the musically trained and the musically untrained group. Similarly, task order did not prove to be influential for acceleration,  $F(1, 34) < 1$ ,  $r = .06$ , impulsiveness,  $F(1, 34) < 1$ ,  $r = .03$ , jerk,  $F(1, 34) < 1$ ,  $r = .07$ , smoothness error,  $F(1, 34) < 1$ ,  $r = .08$ , directness index,  $F(1, 34) < 1$ ,  $r = .14$ , size,  $F(1, 34) < 1$ ,  $r = .04$ , or height,  $F(1, 34) = 1.05$ ,  $r = .17$ . We observed no significant interaction effects between the musical style and musical background. The results were as follows: for acceleration,  $F(1, 34) = 1.80$ ,  $r = .22$ ; for impulsiveness,  $F(1, 34) = 1.18$ ,  $r = .18$ ; for jerk,  $F(1, 34) = 2.08$ ,  $r = .24$ ; for smoothness error,  $F(1, 34) = 2.08$ ,  $r = .24$ ; for directness index,  $F(1, 34) = 2.17$ ,  $r = .24$ ; for size,  $F(1, 34) < 1$ ,  $r = .16$ ; for height,  $F(1, 34) < 1$ ,  $r = .04$ . In addition, there was no significant interaction effect between the musical style and the task order for

TABLE 2. Descriptive Data (Mean and Standard Deviation) for the Semantic Differential Scales, per Group and Musical Style.

	Heroic style		Lyric style	
	Musically trained	Musically untrained	Musically trained	Musically untrained
heavy - light	0.76 (0.73)	1.02 (1.00)	4.02 (1.96)	3.91 (2.03)
vigorous - frail	1.83 (0.82)	1.96 (1.18)	5.74 (1.77)	5.98 (1.54)
rough - delicate	2.69 (1.27)	2.98 (1.65)	7.37 (1.35)	7.28 (1.11)
hard - soft	2.30 (1.06)	2.65 (1.31)	7.35 (1.10)	7.22 (1.25)
fast - slow	4.20 (2.18)	3.20 (1.75)	6.89 (1.70)	6.81 (1.53)
nervous - tranquil	1.07 (1.24)	1.28 (1.43)	4.31 (1.82)	4.56 (1.85)
energetic - soothing	0.96 (1.54)	0.72 (0.56)	4.70 (1.80)	4.89 (1.72)
active - passive	0.69 (0.61)	1.00 (0.82)	3.59 (1.72)	3.94 (1.74)
rigid - fluent	3.02 (2.23)	3.00 (2.11)	5.50 (1.40)	4.69 (1.78)
anxious - secure	2.70 (1.49)	2.8 (1.35)	5.74 (2.01)	5.54 (2.45)
worried - carefree	1.06 (1.16)	0.81 (0.87)	2.91 (2.02)	3.22 (2.25)
serious - playful	2.31 (1.59)	2.33 (1.13)	4.74 (2.24)	4.74 (2.06)
compact - airy	2.26 (1.81)	2.59 (1.28)	4.02 (1.70)	3.46 (1.11)
dense - diffuse	1.37 (1.34)	2.22 (1.28)	4.33 (1.40)	3.54 (1.48)
regular - chaotic	3.91 (2.17)	3.67 (2.07)	1.28 (1.12)	1.46 (1.44)
balanced - unbalanced	5.56 (2.31)	5.35 (1.87)	4.03 (2.20)	4.00 (2.12)
big - small	2.22 (1.77)	2.50 (1.19)	5.74 (1.93)	5.87 (1.64)
broad - narrow	1.50 (1.82)	1.87 (1.32)	2.91 (1.81)	3.35 (1.56)
thick - thin	1.56 (1.62)	2.02 (1.3)	4.59 (1.67)	4.61 (1.35)
exalting - serene	1.20 (1.31)	1.00 (0.78)	4.96 (1.53)	4.78 (1.86)
high - low	3.70 (2.18)	4.11 (1.83)	4.28 (1.61)	4.80 (1.79)
vertical - horizontal	5.09 (2.67)	4.33 (1.37)	5.83 (2.06)	5.59 (1.41)
dominant - humble	0.81 (1.08)	0.89 (0.63)	4.39 (1.72)	4.78 (1.33)
bright - muted	2.07 (1.74)	1.96 (1.54)	2.17 (1.97)	2.30 (1.74)

acceleration,  $F(1, 34) = 2.34$ ,  $r = .22$ , impulsiveness,  $F(1, 34) = 1.18$ ,  $r = .18$ , jerk,  $F(1, 34) = 2.62$ ,  $r = .27$ , smoothness error,  $F(1, 34) < 1$ ,  $r = .16$ , directness index,  $F(1, 34) = 3.06$ ,  $r = .29$ , size,  $F(1, 34) < 1$ ,  $r = .03$ , or height,  $F(1, 34) < 1$ ,  $r = .10$ . Finally, we found no significant interaction between musical style, musical background and task order for acceleration,  $F(1, 34) < 1$ ,  $r = .16$ , impulsiveness,  $F(1, 34) < 1$ ,  $r = .02$ , jerk,  $F(1, 34) < 1$ ,  $r = .15$ , smoothness error,  $F(1, 34) < 1$ ,  $r = .01$ , directness index,  $F(1, 34) < 1$ ,  $r = .13$ , size,  $F(1, 34) < 1$ ,  $r = .05$ , or height,  $F(1, 34) < 1$ ,  $r = .01$ .

#### SELF-REPORT TASK: SEMANTIC DIFFERENTIAL SCALE ANALYSIS

The data from the participants' responses on the semantic differential scales are shown in Table 2, organized according to group (musically trained/untrained) and musical style (heroic/lyric). A three-mode principal component analysis (PCA) was performed in order to reveal the relationships between the participants' musical backgrounds, the style of the musical fragments and their responses on the 24 semantic differential scales (Kroonenberg, 1985, 2008; Smilde, Bro, Geladi, & Wiley, 2004; Tucker, 1966). A standard, two-mode PCA analysis would have been inadequate, as an average of the results for one of the modes is required. Thus, there would have

been an implicit assumption that all participants have the same view on the relationship between the music fragments and the Likert scales. A three-mode PCA, however, considers participants as a linear combination of idealized participants, the adjectives as composed of latent adjectives, and the conditions as composed of prototype conditions. Although each can be studied in a separate PCA, analyzing the same variance in the data in three different ways, it is clear that the components (here called, "idealized," "latent," or "prototype" for the three modes) extracted are somehow related. The three-mode PCA does this by considering the data as combinations of mixtures of idealized participants, latent adjectives, and prototype conditions. Thus, using three-mode PCA, it is possible to unravel underlying components in all three modes, to investigate the relationships between them, and to assess possible individual differences (for a more detailed account of the technical aspects, the reader is referred to the Appendix). A  $2 \times 2 \times 2$  model (fragments  $\times$  scales  $\times$  participants) with two components for each of the modes was constructed, which had a fit of 62% to the data, which is more than adequate for interpretation. The first components of each of the three modes accounted for 61%, 58%, and 59% of the variance in the data respectively, the second components for 1%, 4%, and

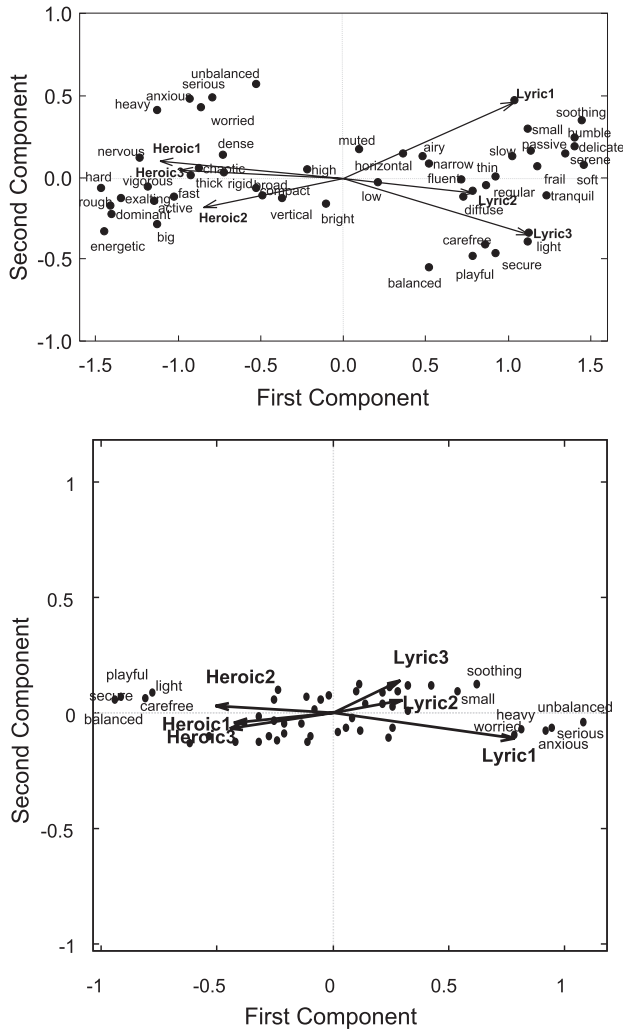


FIGURE 3. Joint biplot for music fragments and scales representing the consensus (top) and the qualitative individual differences (bottom).

3%. For the purposes of this experiment, we were interested in (1) the participants' global consensus about the characterization of the musical fragments in terms of the semantic differential scales, and (2) the individual differences with respect to this consensus characterization which is expressed via the second component of the participants. The high consensus among the participants (61% variance accounted for) dominates the perceived relationships between the scales and the fragments. However, as indicated below, there are different views for a limited number of participants with respect to some scale-fragment combinations which are worth to pursue even though they only represent 1% variance accounted for. For instance, a possible difference between musically trained and untrained participants can be evaluated in this way.

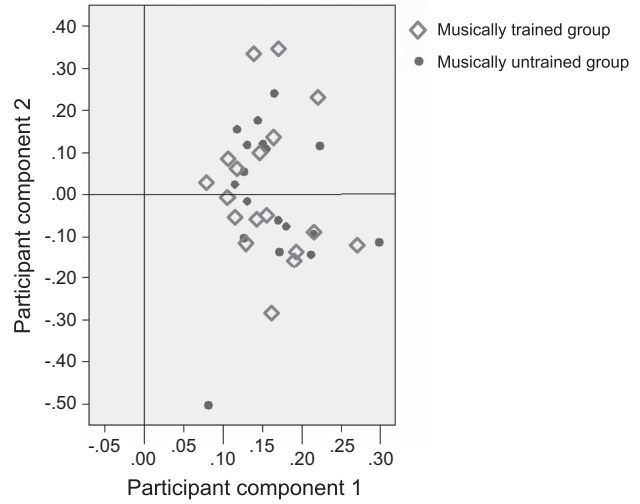


FIGURE 4. Visualization of the subject space.

#### CONSENSUS AMONG PARTICIPANTS

By investigating the component coefficients of the music fragments component jointly with those of the semantic differential scales, the consensus among participants as to the character of the musical fragments can be identified. Both can be projected into a single space, known as a joint biplot representation (Murakami & Kroonenberg, 2003, p. 260) or (Kroonenberg, 2008, p. 273). In three-mode PCA, such joint biplots can be constructed for each component of the third mode (here, participants). Figure 3 (top) shows the participants' consensus regarding the characterization of the musical fragments in terms of the semantic differential scales. The graph shows clearly that the participants rated the heroic and lyric fragments at opposite ends of the scale. Further, nearly all the semantic differential scales show a stark contrast between the heroic and lyric fragments as they project appropriately on the first axis as do the fragments. The extent to which each participant agreed with this consensus judgement of the relations between fragments and scales can be gleaned from the size of the coefficients on the first participant component. Figure 4 shows that all participants gave positive scores to the first component and thus the consensus configuration corresponds with their judgements to some degree. Alternatively, one might say that the configuration is larger for the participants with a large coefficient on the first component and smaller for those with a smaller coefficient. Moreover, by looking at Figure 4, it seems almost self-evident that there is no systematic grouping of the subjects on the basis of their level of music training.



## INDIVIDUAL DIFFERENCES BETWEEN PARTICIPANTS

Figure 3 (bottom) shows which fragments received substantially different judgments from different participants. Whereas the top part of Figure 3 shows the general consensus among the participants, the bottom part also shows those fragments for which participants had mostly small deviations from the consensus, which were usually small. Whether or not a participant deviated from the general consensus is indicated by their scores on the second subject component in Figure 4: a score of zero implies that the consensus represents the individual's judgements; a sizeable positive score for the second participant component (the two most extreme are marked with a "U") implies that they consider the lyric fragments more *unbalanced*, *serious*, and *anxious* than is indicated in Figure 3 (top) and the heroic fragments more *balanced*, *secure*, and *playful*. Thus, their deviation from the general consensus takes the form represented in Figure 3 (bottom). Conversely, if a participant has a negative score on the second subject component (as per the participant marked "T") the situation is reversed: the lyric fragments are considered more *balanced*, *secure*, and *playful* and the heroic fragments more *unbalanced*, *serious*, and *anxious* by that participant than one would derive from the consensus.

## CORRELATION ANALYSIS

An important part of the study was to investigate whether body movements and the expressive concepts associated with the music correlate in terms of expressiveness. For this purpose, we introduced an expressive model based on Laban's Effort/Shape model. In accordance with the results described above, the linguistic concepts and movement characteristics that were associated with the heroic musical fragments are highlighted in grey in Table 1, while those not highlighted are associated with the lyric musical fragments. Arguably, the polarities of movement features identified in our adaptation of the Laban framework correlate consistently with those of the semantic differential scales attributed to the respective music styles. Thus, our qualitative approach brings to light a correlation between movement characteristics and the linguistic labels associated with music.

To quantify this correlation further, we applied a PCA on 30 items (7 movement features and 23 metaphors) with promax rotation. One metaphor – "bright" – was removed from the analysis as the KMO sampling adequacy value for this item was .35, while Kaiser (1974) recommends a bare minimum of .05. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis,  $KMO = .89$ , and all KMO values for individual items were higher than .58. Bartlett's test of

sphericity,  $\chi^2(43) = 3042.83$ ,  $p < .001$ , indicated that correlations between items were sufficiently large for PCA. We ran an initial analysis to obtain eigenvalues for each component in the data. Five components had eigenvalues over Kaiser's criterion of 1, and in combination these explained 80.78% of the variance. However, 57.54% of the variance was due to the first component, which consisted of the majority of the movement features and metaphors, while the other four components only contributed to a minor extent to the total variance. Therefore, we are justified in saying that the movement features and metaphors are correlated.

## Discussion

The results of the experiment show that the participants exhibited contrasting body movement responses corresponding to the heroic and lyric musical fragments: the heroic music evoked body movements that are higher in acceleration, jerk, and impulsiveness, lower in smoothness and directness, larger in size, and more elevated than those performed in response to the lyric music. We observed a significant effect of musical background (i.e., musically trained/untrained) on the size and height of participants' movements. It is possible that the nonmusicians felt more constrained moving freely in response to the music than the musically trained participants. In our exit questionnaire, the participants were asked about how spontaneous they felt while moving to the music. They gave ratings on a five-point Likert scale, with 1 as *totally disagree*, 3 as *neutral*, and 5 as *totally agree*. The median for the musically trained group was 4, while that for the musically untrained group was 2. It seems plausible, therefore, that nonmusicians' less-pronounced, smaller, and less elevated body movements might be put down to their having felt less spontaneous during the motor-tuning part of the experiment. The movement analysis shows that there is no significant interaction effect imposed by the participants' musical background, as the contrast in movement responses to the heroic and the lyric musical fragments is manifest similarly in both in musicians and nonmusicians. This confirms findings of other behavioral studies, which show that people can translate acoustic properties of sound and music into body movements quite consistently (Caramiaux, Bevilacqua, & Schnell, 2010; Godøy, 2010; Küssner, 2013; Kozak, Nymoen, & Godøy, 2012; Leman, Desmet, Styns, Van Noorden, & Moelants, 2009). However, Küssner (2013) reports that musicians are more consistent (i.e., less diverse) in visualizing sound and music by means of drawings. In our study, rather general

measures of expressive movement behavior (up to the level of individual gesture segments) were taken into account, and thereby, data captured over large time scales were reduced to single-value measures. Thus, specific dynamic processes over time were cancelled out. Functional data analysis (FDA) could provide a valuable alternative as it can model data, drawn from continuous processes, as functions of time. This approach enables us to reveal underlying dynamic processes and could provide the means for further investigation of the consistency and diversity of movement behavior in response to music.

The results of the analysis of the semantic differential scales show that the heroic and lyric fragments were associated with opposite adjectives in terms of expressive quality. Moreover, a high degree of consensus among the participants' responses was found, with no differentiation between musicians and nonmusicians. This finding could be related to the work of Bigand and colleagues, who report that emotional responses to music are stable and only weakly influenced by musical expertise (Bigand & Poulin-Charronnat, 2006; Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005). Possibly, this can be explained by the fact that nowadays people are exposed to music on a regular basis, often similar to that used in the experiment, in various everyday-life contexts (film, television, etc.). This allows people to develop a specific attitude towards expression of the music.

Currently, there are very few studies explicitly addressing how body movements performed in response to music might be related to people's descriptions of musical expressivity and emotion (Maes et al., 2010; Sievers, Polansky, Casey, & Wheatley, 2013). For that purpose, we introduced a dimensional model, enabling us to compare body movement features and linguistic labels based on their underlying expressive qualities. These qualities were defined in terms of the taxonomy outlined in Laban's Effort/Shape model. As explained above, the polarities of the movement features correlate consistently with those of the semantic differential scales, in relation the different musical styles. Considering the semantic differential scales, the model was strongly confirmed for the categories Effort-weight, Effort-time, Effort-flow, and Shape-spreading/enclosing. For the category Shape-rising/descending, we observed a tendency towards correlation although substantially weaker. The category Effort-space resulted in negative correlations for the linguistic labels *compact* and *dense*. It could be argued that neither of these labels were perfectly suited for the category Effort-space and that this resulted in negative correlations. However, the overall results indicate that,

generally speaking, body movements performed in response to the music reflect expressive qualities similar to the linguistic labels that are associated with that music. In other words, body movements correlate with the linguistic labels on the level of qualities related to expressive gestures. There is, therefore, a strong basis for supporting our hypothesis that linguistic descriptions of music share common grounds with the body movements that are associated with music. A grounded, modal view on cognition and emotion could possibly explain the connection between body movements in response to music and cross-domain associations of music explicated in linguistic descriptions (Barsalou, 2008; Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012). Studies have shown that sensory, motor (Bangert et al., 2006; Chen, Penhune, & Zatorre, 2008; Haueisen & Knösche, 2001), and introspective (Alluri et al., 2011; Chapin, Jantzen, Kelso, Steinberg, & Large, 2010) brain areas become activated when people listen to music passively. Studies have also indicated that the processing of action-related concepts, abstract concepts, as well as those relating to abstract emotional states recruit neural cell assemblies distributed over sensory (vision, audition, etc.), motor (movement, proprioception, etc.) and introspective (mental states that include affect, motivation, intentions, metacognition, etc.) brain areas (e.g., Aziz-Zadeh & Damasio, 2008; Citron, 2012; Moseley, Carota, Hauk, Mohr, & Pulvermuller, 2012). So, both listening to music, and the processing of concepts create, what Barsalou (1999) calls, *perceptual symbols* (i.e., neural networks distributed over sensory, motor and introspective brain areas). It seems plausible therefore, that if both perceptual symbols partially overlap, that one gets associated with the other. Our study suggests that this overlap can occur, at least partly, in motor-related processing areas.

That the general physical appearance of the movement responses to the music shared similar characteristics across participant groups is a highly interesting finding, all the more because the body movements can be related to a vocabulary of expressive qualities and concepts. This indicates that the human body can function as a nonverbal means for communicating social meaning, as well as emotion and intentionality in general. Further, it provides support for the idea that the human body functions as a mediator, turning purely subjective phenomena (like feelings, intentions and ideas) into physical phenomena (vision, audition and so on) and vice versa. Moreover, our results suggest that the human body plays an important role in the process of musical meaning formation, albeit not an exclusive one. The present study shows that people's engagement with the music was structured around a dynamic

contrast and tension between opposite qualities that could be related to expressive gestures. For the purpose of the study, the beginning of the first movement of Brahms' *First Piano Concerto* was ideally suited as it covers all of the categories, as well as the oppositions within each category, of Laban's Effort/Shape model. However, the method we have introduced, combining the Effort/Shape model, semantic differential scales, and the various algorithms for the extraction of expressive movement features, could easily be extended to other music as well. Moreover, it would be interesting to investigate whether the Effort/Shape model could be used to annotate people's own descriptions and imagery related to how they perceive musical expressivity.

Based on our findings, it is tempting to conclude that our embodied understanding of music (i.e., action) mediates the cross-domain linguistic descriptions ascribed to the music (i.e., perception). However, we can only draw limited conclusions about the relationship between action and perception. That a correlation exists between both does not imply that there is a causal relationship. However, previous literature claims that the use of metaphors is grounded in, and structured by, physically embodied experiences (Crawford, 2009; Gibbs et al., 2004), and in turn, this idea suggests that moving to music may shape people's interpretation and understanding of its expressiveness. A recent study by Maes and Leman (2013) seems to support this idea. In this study, two groups of children were trained (conditioned) to perform either a happy or a sad choreography in response to the same music, which had an expressively ambiguous character. Afterwards, the children's perception of musical expressiveness was assessed in terms of valence and arousal. The results suggested that the expressive qualities of the choreographed movements had a significant impact on how the children perceived musical expressiveness. In addition, studies have started to devote attention to the effects that people's body movements may exert on how they perceive other aspects of music, like pitch (Repp & Knoblich, 2009), meter (Phillips-Silver & Trainor, 2005, 2007), and musical preference (Sedlmeier, Weigelt, & Walther, 2011). Alongside the results of the present study, it seems

increasingly evident that human motor processes are a substantial - although not exclusive - part of people's engagement with music. Moreover, these results provide an argument for the development of methodologies in music education that exploit an active engagement of people's bodies with music.

## Conclusion

In this experiment, participants were obliged to fulfil two tasks. First, they were to translate the musical expressiveness they experienced into corresponding body movements. Second, they had to rate the perceived association between the music and specific linguistic labels. The results of the study show that contrasts in musical style (heroic/lyric) were clearly reflected in the free body movements made in response to the music, and that participants associated distinct sets of linguistic labels to those musical styles. The results for both tasks were independent of the musical background of the participants (i.e., musically trained/untrained). Moreover, it was shown that there are correlations between body movement features (i.e., action) and linguistic labels (i.e., perception) based on their underlying expressive qualities. This action-perception correlation can possibly be explained by a grounded, modal view on music cognition and perception. However, more research is required to reveal causal relationships between action and perception in musical signification processes.

## Author Note

This work is part of the EmcoMetecca project supported by the Flemish Government (<http://www.ipem.ugent.be/EmcoMetecca>). The authors want to thank Ivan Schepers for the technical support, and Dr. Jon Hargreaves for the grammatical corrections.

*Correspondence concerning this article should be addressed to Pieter-Jan Maes, Department of Musicology, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium. E-mail: pieterjan.maes@UGent.be*

## References

- ALLURI, V., TOIVIAINEN, P., JÄÄSKELÄINEN, I., GLERAN, E., SAMS, M., & BRATTICO, E. (2011). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage*, 59(4), 3677-3689.
- AZIZ-ZADEH, L., & DAMASIO, A. (2008). Embodied semantics for actions: Findings from functional brain imaging. *Journal of Physiology*, 102(1-3), 35-39.

- BANGERT, M., PESCHEL, T., SCHLAUG, G., ROTTE, M., DRESCHER, D., HINRICH, H., HEINZE, H.-J., & ALTENMULLER, E. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *Neuroimage*, 30(3), 917-926.
- BARSALOU, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577-660.
- BARSALOU, L. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645.
- BIGAND, E., & POULIN-CHARRONNAT, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal music training. *Cognition*, 100(1), 100-130.
- BIGAND, E., VIEILLARD, S., MADURELL, F., MAROZEAU, J., & DACQUET, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion*, 19(8), 1113-1139.
- BULLOCK, J. (2007). Libxtract: A lightweight library for audio feature extraction. In K. K. Jensen (Ed.), *Proceedings of the International Computer Music Conference (ICMC)* (pp. 25-28), Copenhagen, Denmark: Re: New – Digital Arts Forum.
- BULLOCK, J. (2008). *Implementing audio feature extraction: implementing audio feature extraction in live electronic music* (Unpublished doctoral dissertation). Birmingham City University, Birmingham, England.
- CAMURRI, A., MAZZARINO, B., RICCHETTI, M., TIMMERS, R., & VOLPE, G. (2004). Multimodal analysis of expressive gesture in music and dance performances. In A. Camurri & G. Volpe (Eds.), *Gesture-based communication in human-computer interaction (LNCS)* (Vol. 2915, pp. 357-358). Berlin, Germany: Springer-Verlag.
- CANNAM, C., LANDONE, C., & SANDLER, M. (2010). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. In A. del Bimbo, S.-F. Chang, & A. Smeulders (Eds.), *Proceedings of the International Conference on Multimedia (ICM)* (pp. 1467-1468). Firenze, Italy. New York, NY: ACM.
- CARAMIAUX, B., BEVILACQUA, F., & SCHNELL, N. (2010). Towards a gesture-sound cross-modal analysis. In S. Kopp & I. Wachsmuth (Eds.), *Gesture in embodied communication and human-computer interaction (LNCS)* (Vol. 5934, pp. 158-170). Berlin, Germany: Springer-Verlag.
- CHAPIN, H., JANTZEN, K., KELSO, J., STEINBERG, F., & LARGE, E. (2010). Dynamic emotional and neural responses to music depend on performance expression and listener experience. *PLoS ONE*, 5(12), e13812.
- CHEN, J., PENHUNE, V., & ZATORRE, R. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex*, 18(12), 2844-2854.
- CITRON, F. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language*, 122(3), 211-226.
- COX, A. (2011). Embodying music: Principles of the mimetic hypothesis. *Music Theory Online*, 17(2), 1-24.
- CRAWFORD, L. (2009). Conceptual metaphors of affect. *Emotion Review*, 1(2), 129-139.
- DENISSEN, J., GEENEN, R., VAN AKEN, M., GOSLING, S., & POTTER, J. (2008). Development and validation of a Dutch translation of the Big Five Inventory (BFI). *Journal of Personality Assessment*, 90(2), 152-157.
- FUJIHARA, T., & TAGASHIRA, N. (1984). A multidimensional scaling of classical music perception. *Japanese Journal of Psychology*, 55(2), 75-79.
- GABRIEL, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58(3), 453-467.
- GIBBS, R., COSTA LIMA, P., & FRANCOZO, E. (2004). Metaphor is grounded in embodied experience. *Journal of Pragmatics*, 36(7), 1189-1210.
- GODØY, R. (2003). Motor-mimetic music cognition. *Leonardo*, 36(4), 317-319.
- GODØY, R. (2010). Gestural affordances of musical sound. In R. Godøy & M. Leman (Eds.), *Musical gestures: Sound, movement, and meaning*. New York: Routledge.
- GUCK, M. (1994). Analytical fictions. *Music Theory Spectrum*, 16(2), 217-230.
- HAEUSEN, J., & KNÖSCHE, T. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience*, 13(6), 786-792.
- HEVNER, K. (1936). Experimental studies of the elements of expression in music. *The American Journal of Psychology*, 48(2), 246-268.
- HUBERTY, C., & MORRIS, J. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin*, 105(2), 302-308.
- JOHNSON, M. L., & LARSON, S. (2003). “Something in the way she moves”-metaphors of musical motion. *Metaphor and Symbol*, 18(2), 63-84.
- KAISER, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36.
- KOZAK, M., NYMOEN, K., & GODØY, R. I. (2012). Effects of spectral features of sound on gesture type and timing. In E. Efthimiou, G. Kouroupetroglou, & S.-E. F. (Eds.), *Gesture and sign language in human-computer interaction and embodied communication (LNAI)* (Vol. 7206, pp. 69-80). Berlin, Germany: Springer-Verlag.

- KROONENBERG, P. M. (1983). *Three-mode principal component analysis: Theory and applications*. Leiden, The Netherlands: DSWO Press.
- KROONENBERG, P. (1985). Three-mode principal components analysis of semantic differential data: The case of a triple personality. *Applied Psychological Measurement*, 9(1), 83-94.
- KROONENBERG, P. M. (1995). *Introduction to biplots for G×E tables* (Research report No. 51). Brisbane, Australia: Centre for Statistics, University of Queensland.
- KROONENBERG, P. (2008). *Applied multiway data analysis*. Hoboken, NJ: John Wiley & Sons.
- KRUEGER, J. (2009). Enacting musical experience. *Journal of Consciousness Studies*, 16(2-3), 98-123.
- KÜSSNER, M. B. (2013). Music and shape. *Literary and Linguistic Computing*, 28(3), 472-479.
- LABAN, R. (1947). *Effort*. London, UK: MacDonald & Evans.
- LABAN, R., & ULLMANN, L. (2011). *Choreutics*. Hampshire, UK: Dance Books Ltd. (Original work published 1966)
- LARSON, S. (2012). *Musical forces: Motion, metaphor, and meaning in music*. Bloomington, IN: Indiana University Press.
- LEMAN, M. (2007). *Embodied music cognition and mediation technology*. Cambridge, MA: MIT Press.
- LEMAN, M., DESMET, F., STYNS, F., VAN NOORDEN, L., & MOELANTS, D. (2009). Sharing musical expression through embodied listening: a case study based on Chinese Guqin music. *Music Perception*, 26(3), 263-278.
- LEVINSON, J. (2006). *Contemplating art: Essays in aesthetics*. New York: Oxford University Press.
- LINDQUIST, K., WAGER, T., KOBER, H., BLISS-MOREAU, E., & BARRETT, L. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, 35(3), 121-202.
- MAES, P.-J., & LEMAN, M. (2013). The influence of body movements on children's perception of music with an ambiguous expressive character. *PLoS ONE*, 8(1), e54682.
- MAES, P.-J., LEMAN, M., LESAFFRE, M., DEMEY, M., & MOELANTS, D. (2010). From expressive gesture to sound: The development of an embodied mapping trajectory inside a musical interface. *Journal on Multimodal User Interfaces*, 3(1-2), 67-78.
- MAUS, F. (1988). Music as drama. *Music Theory Spectrum*, 10, 56-73.
- MEAD, A. (1999). Bodily hearing: Physiological metaphors and musical understanding. *Journal of Music Theory*, 43(1), 1-19.
- MOSELEY, R., CAROTA, F., HAUKE, O., MOHR, B., & PULVERMULLER, F. (2012). A role for the motor system in binding abstract emotional meaning. *Cerebral Cortex*, 22(7), 1634-1647.
- MURAKAMI, T., & KROONENBERG, P. (2003). Three-mode models and individual differences in semantic differential data. *Multivariate Behavioral Research*, 38(2), 247-283.
- NIELZÉN, S., & CESAREC, Z. (1981). On the perception of emotional meaning in music. *Psychology of Music*, 9(2), 17-31.
- O'BRIANT, M., & WILBANKS, W. (1978). The effect of context on the perception of music. *Bulletin of the Psychonomic Society*, 12(6), 441-441.
- OSGOOD, C., SUCI, G., & TANNENBAUM, P. (1957). *The measurement of meaning* (Vol. 47). Urbana, IL: University of Illinois Press.
- PARK, I., & SCHUTZ, R. (2006). Repeated measures and longitudinal data analysis. In T. Wood & W. Zhu (Eds.), *Measurement theory and practice in kinesiology*. Champaign, IL: Human Kinetics Publishers.
- PETERSEN, D. (2008). Space, time, weight, and flow: Suggestions for enhancing assessment of creative movement. *Physical Education and Sport Pedagogy*, 13(2), 191-198.
- PHILLIPS-SILVER, J., & TRAINOR, L. J. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science*, 308(5727), 1430.
- PHILLIPS-SILVER, J., & TRAINOR, L. J. (2007). Hearing what the body feels: Auditory encoding and rhythmic movement. *Cognition*, 105, 533-546.
- REPP, B. H., & KNOBLICH, G. (2009). Performed or observed keyboard actions affect pianists' judgments of relative pitch. *Quarterly Journal of Experimental Psychology*, 62(11), 2156-2170.
- SEDLMEIER, P., WEIGELT, O., & WALTHER, E. (2011). Music is in the muscle: How embodied cognition may influence music preferences. *Music Perception*, 28, 297-306.
- SENJU, M., & OHGUSHI, K. (1987). How are the player's ideas conveyed to the audience? *Music Perception*, 4, 311-323.
- SIEVERS, B., POLANSKY, L., CASEY, M., & WHEATLEY, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences*, 110(1), 70-75.
- SMILDE, A., BRO, R., GELADI, P., & WILEY, J. (2004). *Multi-way analysis with applications in the chemical sciences* (Vol. 978). Hoboken, NJ: Wiley Online Library.
- SWANWICK, K. (1973). Musical cognition and aesthetic response. *Bulletin of the British Psychological Society*, 1(2), 7-13.
- TODOROV, E., & JORDAN, M. (1998). Smoothness maximization along a predefined path accurately predicts the speed profiles of complex arm movements. *Journal of Neurophysiology*, 80(2), 696-714.
- TUCKER, L. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3), 279-311.
- VAN DYCK, E., MAES, P. J., HARGREAVES, J., LESAFFRE, M., & LEMAN, M. (2013). Expressing induced emotions through free dance movement. *Journal of Nonverbal Behavior*, 37(3), 175-190.
- WHITE, W., & BUTLER, J. (1968). Classifying meaning in contemporary music. *The Journal of Psychology*, 70, 261-266.

## Appendix

### DATA STRUCTURE AND RESEARCH QUESTIONS

The semantic differential data described in the main part of the paper have the form of music fragments  $\times$  (semantic differential) scales  $\times$  participants. In other words participants listened to music fragments and used the nine-point bipolar scales (for details of the scales, see Table A1) to indicate the nature of the music. In particular, the music fragments could be heavy versus light, soft versus loud, etc.

The basic aim was to find out how the participants characterised the music fragments in terms of the scales and whether there were differences between the subjects in general and between persons with and without some music training.

Three-mode analysis of three-mode data was chosen as the appropriate technique for the analysis. The technique used — three-mode principal component analysis — is an extension of principal component analysis to three-mode data; for full details of the technique see Kroonenberg (2008) and Murakami and Kroonenberg (2003) for a comparable use of the technique in musical appreciation. Without going into detail we will present primarily an appropriate way to interpret the information displayed in Figures 3 and 4 in the main body of the paper.

### MEAN DIFFERENCES BETWEEN THE MUSICALLY TRAINED AND UNTRAINED PARTICIPANTS

Before entering into a discussion of the relationships between the music fragments and scales we analyzed whether serious mean differences existed between the musically trained and untrained participants. Even though their number is actually much too small for comprehensive multivariate testing and for adequate power, we nevertheless performed per fragment a Hotelling's  $T^2$  test for testing the differences between the groups using all 24 scales as the multivariate response. Multivariate tests provided the following  $p$  values for the six fragments, respectively:  $p = .36, .76, .66, .62, .15, .69$ . In addition, of the 24 scales per fragment, at most two of them showed a univariate significant difference between the two groups with only one scale being significantly different in two different fragments. Given the outcomes of these tests, we take the point of view that it is a reasonable to assume that the observed differences are primarily due to chance, so that we may treat the two groups together. Note that this does not automatically imply that also their perceived relationships between the fragments and scales are negligible.

### NO INDIVIDUAL DIFFERENCES IN STRUCTURE

To start on familiar ground, assume that all participants agree in an equal manner on the structure in the relations between the music fragments and scales. In such a case there is no necessity to look for individual differences and one might as well sum over participants to get a music fragments  $\times$  scales matrix with means, here indicated by  $\mathbf{M}$ . One may consider the mean music fragments  $\times$  scales matrix as the consensus containing the common opinion of the participants about the relations between the scales and the music fragments.

Taking the mean over all participants to create a mean matrix  $\mathbf{M}$  is equivalent to giving each subject a weight  $\frac{1}{K}$ , where  $K$  is the number of participants. If we indicate an arbitrary subject by  $k$  and the fragments by scales data matrix produced by the participant  $\mathbf{D}_k$  then the summation equation looks like

$$M = \frac{1}{K}D_1 + \frac{1}{K}D_2 \dots \frac{1}{K}D_K = \sum_k^K \frac{1}{K}D_k \quad (1)$$

Another way of looking at the situation is to assume that each data matrix can be modelled by the mean matrix plus random error. To describe each data matrix it suffices to describe the mean matrix or consensus as the difference between data matrix and the model matrix or mean is only random error.

A simple analogy would be that we first calculate the mean height of a group of persons and use that mean to estimate or model the height of an arbitrary person from that group. We know that we are not exactly correct but it is the best we can do without any further information.

### DIFFERENCES IN SIZE

It need not necessarily be true that each participant contributes equally to the consensus. Thus let us associate a weight  $w_k$  with each participant  $k$  where the weight indicates the size of the contribution of the participant to the consensus matrix. In the previous paragraph this weight was  $\frac{1}{K}$  for each participant but now the weights are unequal.

$$\bar{M} = w_1D_1 + w_2D_2 \dots w_KD_K = \sum_k^K w_kD_k \quad (2)$$

To take again the inverse look, we can say that each data matrix  $\mathbf{D}_k$  can be modelled by  $\frac{1}{w_k}M$ . Now the mean matrix is no longer sufficient for modelling a participant's data matrix but we have to take into account that some participants have larger  $\frac{1}{w_k} > 0$  configurations than the mean, while others have smaller  $\frac{1}{w_k} < 0$  configurations. Participants who answer randomly have  $\frac{1}{w_k} = 0$ .

TABLE A1. Scale means for music fragments.

High score end		Music Fragments						Low score end	Category
		Heroic 1	Heroic 2	Heroic 3	Lyric 1	Lyric 2	Lyric 3		
dominant	23	<b>8.1</b>	<b>7.8</b>	<b>7.6</b>	3.4	3.0	3.9	humble	6
active	8	<b>8.1</b>	<b>7.7</b>	<b>7.7</b>	3.9	4.1	4.9	passive	2
worried	11	<b>8.0</b>	<b>7.5</b>	<b>7.4</b>	4.5	5.3	5.1	carefree	3
heavy	1	<b>7.9</b>	<b>7.6</b>	<b>7.3</b>	3.5	4.2	4.6	light	1
energetic	7	<b>7.9</b>	<b>7.8</b>	<b>7.6</b>	2.9	2.9	3.9	easeful	2
nervous	6	<b>7.6</b>	<b>7.2</b>	<b>7.6</b>	3.3	3.5	4.0	tranquil	2
exalting	20	<b>7.5</b>	<b>7.7</b>	<b>7.0</b>	2.9	2.9	3.6	serene	5
dense	14	<b>6.9</b>	<b>6.3</b>	<b>6.0</b>	3.7	4.3	4.2	diffuse	4
broad	18	<b>6.8</b>	<b>6.8</b>	<b>6.5</b>	4.4	<b>5.5</b>	5.0	narrow	5
thick	19	<b>6.6</b>	<b>6.6</b>	<b>6.3</b>	3.4	3.3	3.6	thin	5
clear	24	<b>6.1</b>	<b>6.0</b>	<b>6.4</b>	<b>5.8</b>	<b>5.7</b>	<b>6.4</b>	muted	6
unbalanced	16	<b>5.8</b>	<b>5.6</b>	5.0	3.7	4.2	4.1	balanced	4
compact	13	<b>5.7</b>	<b>5.8</b>	<b>5.8</b>	4.4	3.9	4.5	airy	4
rigid	9	5.0	5.2	5.2	3.3	2.2	3.3	fluent	3
low	21	4.4	3.6	3.8	4.8	4.6	4.2	high	6
playful	12	2.0	2.4	2.6	5.1	4.3	4.8	serious	3
small	17	2.0	2.6	2.6	<b>6.0</b>	<b>6.0</b>	5.4	big	5
horizontal	22	4.5	4.5	5.2	<b>5.5</b>	<b>6.1</b>	<b>5.5</b>	vertical	6
frail	2	1.7	2.0	2.0	<b>6.3</b>	<b>5.8</b>	<b>5.5</b>	vigorous	1
secure	10	2.3	2.8	3.2	<b>5.9</b>	5.4	<b>5.6</b>	anxious	3
slow	5	3.8	3.6	3.7	<b>7.1</b>	<b>7.5</b>	<b>5.9</b>	fast	2
regular	15	3.8	4.3	5.0	<b>6.9</b>	<b>7.0</b>	<b>6.7</b>	chaotic	4
Soft	4	2.3	2.3	2.8	<b>7.3</b>	<b>7.6</b>	<b>6.9</b>	hard	1
delicate	3	2.5	2.8	3.2	<b>7.3</b>	<b>7.5</b>	<b>7.2</b>	rough	1

Note: Scales run from 1 to 9; 5 is the neutral category. Means of 5.5 or higher are set in bold. The column "Category" refers to the Effort/Space model categories.

In our length analogy we could try to predict the length of a person on the basis of the mean length over all persons to get a better estimate for the real length. To model the difference in length we multiply the mean length for each participant with some coefficient  $w_k$ .

#### DIFFERENCES IN QUALITY

In practice there are more differences between participants than only differences in size. There are generally also difference in quality. Some participants have a different interpretation or evaluation of the relation between the soft – hard and the light – dark contrasts for particular music fragments. They agree overall with the consensus but for these fragments they consider the two contrasts more appropriate than one could be described by the consensus alone. In component analysis such effects show up as an additional component. This additional component for the participants has an additional matrix  $M_2$  associated with it.

$$\bar{M}_2 = v_1 D_1 + v_2 D_2 \dots v_K D_K = \sum_k^K v_k D_k \quad (3)$$

Thus  $M_2$  is constituted with different weights, where a small weight for a participant means that his or her

data do not contribute to the construction of the additional mean-like matrix.

If we look at the inverse problem of estimating the original scores from the model information we can model the data of a participant  $k$  by adding the mean and the mean-like matrix with different weights. We should choose these weights in such a way that the difference between the data and the weighted sum of the two mean matrices is as small as possible.

In the (imperfect) length analogy in the case when we have men and women, we try to find a coefficient to multiply the mean with and one to apply to the typical deviation, for instance the average deviation over persons. The length of a man  $m1$  is thus  $length_{m1} = \frac{1}{w_{m1}} \times MeanLength + \frac{1}{v_{m1}} \times Typical\ deviation + error_{m1}$  and for a woman  $v1$  this is  $length_{v1} = \frac{1}{w_{v1}} \times MeanLength + \frac{1}{v_{v1}} \times Typical\ deviation + error_{v1}$ , where  $v_{m1}$  is positive and  $v_{v1}$  is negative because on average men are taller than women and thus we predict a higher weight.

#### Plotting Mean Matrices: Biplots

The relevant mean matrix for our study was portrayed in Table A1. It is obvious that such a matrix cannot

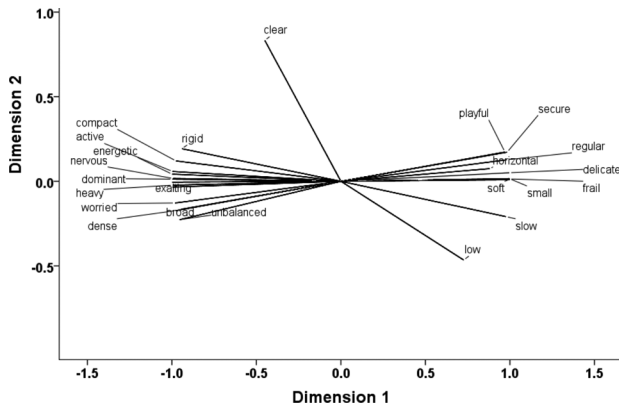


FIGURE A1. Biplot of the mean matrix.

fruitfully be inspected by eye even though we have applied a few tricks to make it as easy as possible. For a proper investigation one generally subtracts the mean of each scale on the assumption that it is not the absolute scale value one is interested in but only how the fragments are located with respect to the mean of each scale or alternatively how the scales apply to a particular fragment. These centred mean scores are used to construct a biplot (Gabriel, 1971).

#### Interpretational Rules for (Joint) Biplots

The basic rules for the interpretation of biplots are the following (see Figure A1). (1) Due to the centering, the origin represents the mean of each of the scales. (2) One of the modes is represented by points (here, scales) and the other mode by arrows (here, fragments). The relation between a music fragment and a scale can be found by dropping a perpendicular from the scales onto the arrow of the fragment. A fragment located in origin has an average score on all scales. If the end point of a scale projects onto the end of a fragment arrow, the fragment has a high value for that adjective (e.g., *small*) and a low value for the adjective associated with the other end of the scale (e.g., *fast*). The relative positions of the scales' projections on a fragment indicate to what extent the participants associated the scales with that fragment. (3) Arrows that have small angles have similar projections on them. Thus with respect to the scales, fragment arrows close together were characterized in the same way. (4) The importance of a scale to characterize a fragment depends only on the position of the scale's projection on the fragment and not on the distance a scale has to the fragment. For instance, a scale above and a scale below the arrow of a fragment can be equally important in describing the fragment even though they are far apart, as long as their projections on that fragment are more or

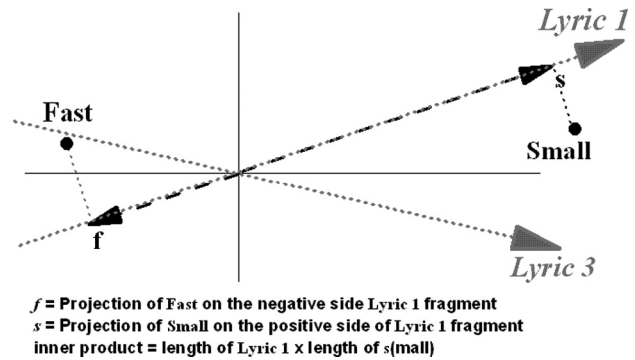


FIGURE A2. Biplot of the mean matrix.

less the same. (5) Scale points close together have more or less the same profile over the fragments. (6) Long fragment arrows indicate that the participant's award scale scores to those fragments far away from the scale mean. When an arrow is short, the participants give the fragment only mean scale scores (little variance).

In technical terms, the aim of using biplots is to portray simultaneously the rows and the columns in such a way that the inner products between the rows and the columns reproduce the values in the centred mean matrix as well as possible. The size of an inner product is the product of the length of the fragment vector from the origin (e.g., Lyric 1) times the length of the projection of the scale on the fragment vector (e.g., the length of  $s$  in the figure).

#### FROM TWO-MODE MEANS TO THREE-MODE JOINT BIPLOTS

A problem with the approach sketched above is that we have to deal with possibly very large tables for which it is difficult to see the structure and to interpret the patterns properly. The standard solution for examining the patterns in such tables is to make a biplot in which we portray both the rows (music fragments) and the columns (scales) in such a way that we can view the patterns from the plot. For the original description of a biplot see Gabriel (1971); in Kroonenberg (1995) an elementary description is given with detailed interpretational guidelines, also for three-mode data discussed here. The biplot is constructed using the same computational algorithm that lies at the heart of principal component analysis.

In Table A1 we have presented the mean or consensus matrix for our data and Figure A2 is the two-dimensional biplot for this mean matrix. It shows a reasonable close resemblance to the joint biplot in the main body of the paper, because in this experiment there was a general consensus about the relations between the fragments and the scales.



We could also produce a biplot for the mean-like matrix  $\mathbf{M}_2$  we discussed in Section “Differences in Quality.” To get a proper model for the original data we have to find a way to combine the two biplots. This is what is done via the three-mode principal component analysis of the complete three-mode dataset.

### THREE-MODE PCA

In order to interpret the outcomes of a three-mode PCA the following steps need to be followed: (1) preprocessing (centring and/or normalizing); (2) selecting model complexity (selecting the number of components for each mode); (3) carrying out the analysis and evaluating its appropriateness; (4) constructing plots for interpretation (in particular, *joint biplots*).

#### Preprocessing

First, the data have to be centered in an appropriate manner. In standard PCA of a participant by variable matrix this is done by calculating deviation scores for the participants with respect to the variable means. In semantic differential data for a single person, the equivalent is to calculate deviation scores for the concepts (here, music fragments) with respect to the scale means. For several persons the procedure is the same, each person’s judgements are analyzed by looking at deviation scores of their own scale means; this process is referred to as *centering*. More formally,  $\tilde{x}_{ijk} = x_{ijk} - \bar{x}_{.jk}$

#### Complexity Selection

The complexity of the three-mode PCA model is determined by the numbers of components of the three modes. On the one hand, one desires sufficient components to describe the data adequately with enough detail for interpretation, on the other hand the more parsimonious the model is the less parameters have to be interpreted. To find an adequate model we have made use of both a deviance plot portraying the deviance or residual sum of squares of the models versus their degrees of freedom and a three-mode scree plot portraying the deviance versus the sums of the numbers of components from all modes of the data array; for a detailed explanation, see e.g., Kroonenberg (2008, Chapter 8). The examination of the two plots led us to choose the  $2 \times 2 \times 2$  model.

### THREE-MODE JOINT BIPLLOTS

Similar to the biplot of the mean matrix one can construct biplots that depict the relationships between the fragments and the scales; they are generally called *joint biplots*. There is a joint biplot for each of the components of the participants. As the first component of the participant space shows all positive coefficients for the

participants, it implies that all subjects weight the associated joint biplot in the same manner so that we may speak of a consensus joint biplot. Then the size of the coefficients indicates to what extent this consensus biplot depicts the judgements of the participants (see Table A2). For instance, participant *M2* has a large coefficient on the first component (0.31) so that the consensus contributes a lot towards describing the answers while for participant *M12* this value 0.06 so that the consensus configuration is practically irrelevant

TABLE A2. Participant space sorted with respect to both first and second components.

Id	Sorted		Id	Sorted	
	1st	2nd		1st	2nd
<i>M12</i>	.06	-.37	<i>NF14</i>	.16	<b>.38</b>
<i>NF1</i>	.09	.01	<i>NF12</i>	.12	<b>.32</b>
<i>NF4</i>	.09	.07	<i>M16</i>	.15	<b>.26</b>
<i>NF13</i>	.10	.10	<i>NF10</i>	.22	<b>.26</b>
<i>NF5</i>	.10	-.05	<i>M17</i>	.12	<b>.22</b>
—	—	—	<i>M5</i>	.11	<b>.21</b>
<i>M5</i>	.11	.21	—	—	—
<i>M18</i>	.11	-.01	<i>M8</i>	.21	.16
<i>M10</i>	.11	-.07	<i>NF16</i>	.16	.13
<i>M6</i>	.12	.07	<i>NF9</i>	.15	.12
<i>M14</i>	.12	.10	<i>M1</i>	.17	.10
<i>NF12</i>	.12	.32	<i>M14</i>	.12	.10
<i>M17</i>	.13	.22	<i>NF13</i>	.10	.10
<i>NF3</i>	.13	-.08	—	—	—
<i>M9</i>	.13	.06	<i>NF4</i>	.09	.07
<i>NF2</i>	.14	-.11	<i>M6</i>	.12	.07
<i>M15</i>	.14	-.07	<i>M9</i>	.13	.06
<i>M4</i>	.15	-.04	<i>M13</i>	.23	.02
<i>NF9</i>	.15	.12	<i>NF1</i>	.09	.01
<i>M16</i>	.15	.26	<i>M18</i>	.11	-.01
<i>NF16</i>	.16	.13	<i>NF6</i>	.16	-.03
<i>NF14</i>	.16	.38	<i>M3</i>	.21	-.03
<i>NF6</i>	.16	-.03	<i>M4</i>	.15	-.04
<i>NF7</i>	.17	-.30	<i>M11</i>	.18	-.04
<i>M1</i>	.17	.10	<i>NF5</i>	.10	-.05
<i>NF18</i>	.18	-.10	<i>M10</i>	.11	-.07
<i>M11</i>	.18	-.04	<i>M15</i>	.14	-.07
—	—	—	<i>NF3</i>	.13	-.08
<i>NF8</i>	<b>.20</b>	-.19	<i>NF11</i>	.20	-.08
<i>NF11</i>	<b>.20</b>	-.08	<i>NF17</i>	.24	-.09
<i>M7</i>	<b>.21</b>	-.21	—	—	—
<i>M8</i>	<b>.21</b>	.16	<i>NF18</i>	.18	-.10
<i>M3</i>	<b>.21</b>	-.03	<i>NF2</i>	.14	-.11
<i>NF10</i>	<b>.22</b>	.26	<i>M2</i>	.31	-.18
<i>M13</i>	<b>.23</b>	.02	<i>NF8</i>	.20	-.19
<i>NF17</i>	<b>.24</b>	-.09	<i>NF15</i>	.27	-.19
<i>NF15</i>	<b>.27</b>	-.19	—	—	—
<i>M2</i>	<b>.31</b>	-.18	<i>M7</i>	.21	<b>-.21</b>
			<i>NF7</i>	.17	<b>-.30</b>
			<i>M12</i>	.06	<b>-.37</b>

for this participant. However, *M12* has a value of -0.37 on the second component so that the configuration belonging to this second participant component corresponds well to the individual configuration of participant *M12*. For *M7* the two configurations are equally important so that we have to add them to see what this participant's private configuration is like.

Table A2 also shows the second component for the subjects and it supplies the coefficients with which to multiply the second joint biplot produced by the three-mode analysis, analogously to the mean-like matrix discussed in Section "Differences in Quality."

The patterns in the joint biplots for our semantic differential data and their evaluation is explained in the main body of the paper.

#### SUMMARY

In this Appendix we have tried to explain the background to the analyses presented in the main body of the paper, in particular that the construction of the joint biplots and the rationale behind them. For more details one is advised to read the relevant chapters of Kroonenberg (1983, Section 6.10) and/or Kroonenberg (2008, Section 11.5.3) and the examples in those books.