Article

MUSICÆ SCIENTIÆ

Musicae Scientiae
15(2) 270–287

DOI: 10.1177/1029864911401173
msx.sagepub.com
SAGE

Music's impact on the visual perception of emotional dyadic interactions

Ramona Kaiser
Music Cognition and Action Group, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; Max Planck International Research Network on Aging

Peter E. Keller
Music Cognition and Action Group, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Abstract
Research showing that emotions can be recognized in point-light displays of human dyadic interactions was extended in the current study by investigating the impact of music on the perception of normal and exaggerated expressions of happiness, contentedness, sadness and anger in such visual stimuli. Sixteen musically untrained participants viewed short video clips of these emotion portrayals, each presented with emotionally compatible (e.g., happy music accompanies a happy interaction) and emotionally incompatible piano music (e.g., sad music accompanies a happy interaction). It was hypothesized that music will increase the accuracy of emotion judgements in displays where auditory and visual information is compatible relative to displays with incompatible audio-visual information. A two-dimensional emotion space was used to record participants' judgements of emotion in only the visual stimuli. Results indicated that music affected the accuracy of emotion judgements. Happiness and sadness were perceived more accurately in compatible than in incompatible conditions, while the opposite was the case for contentedness. Anger was perceived accurately in all conditions. Exaggerated expressions of sadness, which were evaluated more accurately than normal expressions of sadness, were also found to be resistant to the music. These findings can be interpreted in the light of previous research demonstrating that music's cross-modal impact depends on the degree of emotional ambiguity in the visual display. More generally, the results demonstrate that the perception of emotions in biological motion can be affected by music.

Keywords
cross-modal perception, emotions, music, nonverbal communication, point-light display

Corresponding author:
R. Kaiser, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1A, 04103 Leipzig, Germany
Email: rkaiser@cbs.mpg.de

Introduction

The expression and perception of emotions have been topics of inquiry for centuries and much is known about the communication of emotion through nonverbal channels such as music and body language. However, given the popularity of multimedia art forms such as opera, musicals and film, surprisingly little is known about the impact of music on the perception of bodily emotion communication. The present study investigates this impact and thereby builds upon research addressing emotion communication in (1) music performance, (2) visual nonverbal behaviour and (3) multimodal displays in animation and film.

Investigations concerning emotion communication in music can be traced back to the early 19th century (Gabrielsson & Juslin, 2003). Much of this work has sought to answer the question of how music can communicate emotions through the analysis of a musical piece's structure as specified in a notated score. The effects of factors such as mode (major vs. minor), rhythm, and their interaction, for example, have been thoroughly investigated in this research tradition (for a review see Gabrielsson & Lindström, 2001). In recent decades, a second stream of research on music and emotion research – focusing on the analysis of musical performances rather than scores – has increased in prominence (see Clarke, 2004; Juslin & Timmers, 2010). Empirical studies in this performance-based tradition have confirmed that skilled performers can communicate different emotions through variations in, for example, articulation, sound intensity level and tempo (e.g., Gabrielsson & Juslin, 1996; Madison, 2000) with accuracy as high as that for verbal or facial stimuli (see Juslin, 2001).

Studies of performers' movements, which are used not only for tone production but also to communicate with co-performers and the audience (see Davidson & Correia, 2002), have revealed that a performer's body language can communicate expressive intensity (e.g., Davidson, 1993; Vines, Krumhansl, Wanderley, & Levitin, 2006) as well as specific emotions (Dahl & Friberg, 2007). Related work on vocal performance has found that singers' facial expressions communicate emotions to observers (e.g., Livingstone, Thompson, & Russo, 2009; Thompson, Russo, & Quinto, 2008).

The ability to communicate emotions in facial and bodily expressions has also been demonstrated in research on visual nonverbal behaviour outside the music domain. For a long time this research was dominated by studies on the ability of facial expressions to communicate emotions (for reviews see Russell, 1994; Ekman, 1993). This dominance may be explained by the prevailing belief (at the time) that the body is only able to express the *quantity* (i.e., intensity) rather than the *quality* of an emotion (for a discussion of that topic see Wallbott, 1998). However, Wallbott (1998) demonstrated by analysing bodily expressions of 14 emotions that bodily motion can communicate both the quantity and quality of emotions, with some movements being specific to certain emotions. Another study demonstrating that bodily motions can communicate emotions was conducted by Atkinson, Dittrich, Gemmell, and Young (2004). They compared emotion perception in pictures and videos, with the pictures being created by selecting images from the videos at the peak of the emotion expression. Although the emotional information was the same for both stimuli (i.e., the emotional expression itself as its peak) the evaluation of the emotions was always more accurate for the moving stimuli than for the static postures. The above-mentioned studies presented normal pictures or videos, which are often referred to as full-light (FL) stimuli, as the person and their environment are presented in a conventional way. Besides this technique, point-light (PL) animations are frequently used to study the perception of movement kinematics.

In the PL technique, a person's actions are filmed with small lights or reflective material attached to his or her body. In the subsequently displayed visual material, only the points, and therefore the mere motion of the body, is visible. An advantage of PL-stimuli is that they isolate the dynamic, kinematic aspect of bodily communication (i.e., without any static form information). Research has shown that people are able to detect a figure's identity and to discriminate between different actions in PL-displays (e.g., Dittrich, 1993; Runeson & Frykholm, 1983; Troje, Westhoff, & Lavrov, 2005). Moreover people are able to perceive a range of emotions correctly for dancing (Dittrich, Troscianko, Lea, & Morgan, 1996) and simple actions such as drinking, knocking and walking (e.g., Chouchourelou, Matsuka, Harber, & Shiffrar, 2006; Pollick, Paterson, Bruderlin, & Sanford, 2001). The evaluation of emotions can be as precise as for FL-stimuli, at least for some emotions (see Atkinson et al., 2004), and is still above chance even with inverted or reversed displays (Atkinson, Tunstall & Dittrich, 2007). Presenting PL-stimuli of normal and exaggerated expressions of anger, disgust, fear, happiness and sadness, Atkinson et al. (2004) found increased evaluation accuracy for the exaggerated PL-displays of all emotions except for sadness.

Although emotional expressions occur most frequently within social contexts (Ekman, 1999; Sloboda & Juslin, 2001), most existing studies use a single-person setting and thus neglect the social aspect of emotions. Focusing on emotional expression in this way may capture the perception of emotions incompletely. To address this issue, Clarke, Bradshaw, Field, Hampson and Rose (2005) investigated the impact of social interaction on the perception of emotions. Analysing the emotions of anger, joy, sadness, disgust, fear and romantic love, they found that all of them, apart from disgust, were recognized correctly from the body movements of interacting couples. While for some emotions an increase in evaluation accuracy was observed with the presence of an interacting partner, the presence of a non-interacting partner (i.e., an actor plus his or her reflection) decreased evaluation accuracy for some emotions. Clarke et al. (2005) claimed that their findings demonstrate the significance of social interaction for the process of emotion communication, with the entire scene being considered when people evaluate the emotional content of a situation. If so, how does acoustical information affect the evaluation of a social scene?

Despite the prominent role of music in film, opera, theatre, games and ritual (see e.g., Dissanayake, 2006; Lipscomb & Tolchinsky, 2005; Zehnder & Lipscomb, 2006), only a handful of studies have investigated the cross-model interaction of visual and auditory emotion communication. In a seminal study, Cohen (1993) demonstrated that the characterization of a simple animation of a bouncing ball is affected by music. When presented in silence, the ball was evaluated as increasingly happy the higher and/or faster it bounced. Combined with a simple melody, varying in pitch height (i.e., low, medium or high) and/or tempo (i.e., slow, moderate or fast), the same stimulus was rated differently. More specifically, a high and/or fast bouncing ball was perceived to be less happy when presented with structurally incompatible music (i.e., low pitched and/or slow music). Conversely, a low and/or slow bouncing ball was evaluated to be less sad when combined with high pitched and/or fast music (Cohen, 1993, 2005). Bolivar, Cohen and Fentress (1994) found similar effects using more complex stimuli. They presented short films of friendly or aggressive social interactions between wolves combined with commercially used jingles that had been previously evaluated as friendly or aggressive. Again the music influenced the rating of the perceived emotions. Emotionally compatible music (e.g., a friendly social interaction presented with friendly sounding music) enhanced the perception of the intended emotional character of the interaction, while incompatible music (e.g., a friendly interaction presented with aggressive sounding music) had a contrary effect.

However, music does not necessarily impact all types of visual stimuli to the same extent. Cohen (1993) presented two musical excerpts, suggestively titled “Conflict” and “Say Hello To

Love," with two short film sequences that varied in degree of emotional ambiguity. While one scene, a fight between two men, was clear in its aggressive meaning, the encounter of a woman and a man in the second scene could be perceived as either romantic or aggressive (Cohen, 2005). Only the evaluation of this second, ambiguous scene was found to be affected by the presence of music. Relative to the film-only ratings this scene was evaluated as more active and less positively valenced when presented with the aggressive music. For the romantic music the contrary effect was found: the scene was perceived to be less active and more positive compared to the film-only ratings, although the impact of the romantic music was less pronounced compared to the aggressive music.

As these studies have shown that music can affect the perception of different visual stimuli, how would music affect the perception of emotions communicated in the kinematics of human body language? Motion is closely linked to speech, as the body contributes important information via nonverbal communication channels. A close link can also be found between music and speech, with many acoustical patterns being used in a similar way to communicate emotions (see Juslin & Laukka, 2003). The consideration of both music and motion therefore raises the possibility of analysing the interplay between two different channels of nonverbal emotion communication. Van den Stock, Peretz, Grèzes and De Gelder (2009) investigated the impact of music on the perception of emotional body language presenting short FL-movies of single actors with blurred faces. Each actor expressed happiness and sadness while drinking (i.e., raising a glass and putting it back on a table). These videos were presented alone or accompanied by emotionally compatible or incompatible classical music. By calculating the proportion of happy responses, it was shown that happy music affects the perception of the happy and sad videos. Both were evaluated as happier than presentations with sad music or videos without music.

The present study focuses on music's impact upon emotional body language expressing happiness, contentedness, sadness and anger. In order to restrict visual emotional cues to kinematic information, PL-stimuli were used instead of FL-videos. To investigate the perception of these cues in a naturalistic, social setting, an interpersonal dialogue – based on that used by Clarke et al. (2005) – was employed when creating these PL-stimuli. Following Atkinson et al. (2004), normal and exaggerated emotional versions of this interpersonal dialogue were produced in order to examine music's impact upon the perception of emotional body language under conditions of varying visual ambiguity.

Emotion perception was measured by requiring participants to respond to the PL-displays using a two-dimensional emotion space (2dES). This emotion space was created by combining the approaches of the two-dimensional emotion space developed by Schubert (e.g., 1996, 2004) and the Affect Grid employed by Russell, Weiss and Mendelsohn (1989), both based on the dimensions of valence and arousal. An additional decoy task was employed to reduce possible evaluation bias. Such evaluation bias could arise if participants gain awareness of the research question. Asking the participants solely to rate the perceived emotion in visual stimuli that are presented with emotionally compatible or incompatible music may lead to such awareness. We assumed that introducing an additional visual task would divert participants' attention from the music to the visual information, thereby obscuring the purpose of the study. Such a visual decoy task was utilized by asking the participants to report whether an earlier presented target video was included among a series of videos that were accompanied by music, in addition to the experimental task of evaluating the perceived emotion in the visual stimuli.

It was hypothesized that – despite the decoy task – music should have an impact on the evaluation of emotion in PL-displays of social interaction, with greater recognition accuracy in the compatible than the incompatible condition. It was further hypothesized that the perception of the exaggerated, less ambiguous, emotion expressions (cf. Atkinson et al., 2004) would be affected less by the music than the normal expressions.

Method

Participants

Eight females and eight males ($N = 16$) with an average age of 24.64 years (range 19–31 years) participated as paid volunteers in the experiment. Six other individuals (average age 23.5; range 19–26 years) participated in a pretest used for stimulus selection (see below). None of the participants had formal music training. All had normal or corrected-to-normal vision and none of the participants reported relevant hearing impairment.[1]

Design

To investigate music's impact on the visual perception of emotions, auditory and visual stimuli were presented in a within-subjects design. Expressions of happiness, contentedness, sadness and anger (Emotion) were presented as PL-stimuli representing each emotion in a normal and exaggerated manner (Display). Each of the PL-stimuli was presented with emotionally compatible and incompatible music (Music) employing three musical pieces (Pieces). In emotionally compatible conditions the same emotion was expressed in the visual and auditory stimuli. In emotionally incompatible conditions, expressions of happiness were crossed with expressions of sadness, and expressions of contentedness were crossed with expressions of anger, yielding four types of incompatible stimuli.

Materials

Video. PL-displays of interacting couples were used as visual stimuli. To record these displays, two female and two male professional actors (age range 29–48 years) with an average working experience of 14.5 years (range 4–30 years) were paid for their services. A dialogue was created by translating and modifying the one used by Clarke et al. (2005) (see Appendix Table A1). The actors were instructed to learn these lines by heart one week before the recording. They were also informed that their task would be to perform this dialogue in a happy, content, sad and angry way, with each emotion being portrayed twice, first normally and then in an extremely exaggerated manner.

The actors were told that each emotion should be expressed using their voices and body movements and that sound and video recordings would be made of each performance (although these were not actually used in the experiment). A Vicon motion capture system was used to record the position of 25 reflective markers attached to each actor with a sampling frequency of 200 frames per second. The actors were instructed to perform the emotions without using symbolic gestures, such as holding their belly for laughter. They were allowed to move freely in the room, but were instructed to avoid physical contact with each other. Each actor worked with every other actor and was randomly assigned as person A or B, with person A always standing on the right hand side from the observer's point of view. The recordings were done during a one-week period, with each pair of actors working separately (i.e., no pair saw the performances of other pairs). Each recording of the normal and the exaggerated

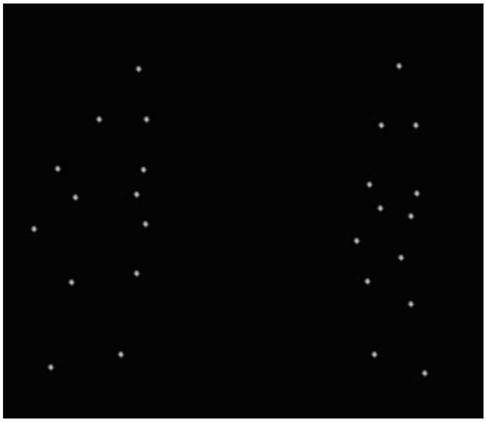

Figure 1. A still image from a PL-display of two actors expressing anger. Each actor is represented by 13 points of light (indicating the head, shoulders, elbows, hands, hips, knees and feet)

emotional performances was repeated until the performance was judged by the actors to be satisfactory (range 1–10 times) in the sense that it was error-free and produced with the best possible interpretation. The first 7s of the best performances of each pair were selected and the recordings were modified in such a way that only 13 markers on each actor were visible in the PL-stimuli presented to the participants (see Figure 1).

Wallbott (1998) reported in his analyses of affective body expressions that, amongst other things, emotional expressions differ most significantly in performed motions and postures of arms and hands. Therefore PL-markers attached to the actors' hands, rather than wrists, were shown in the stimuli (as in Atkinson et al., 2004). To investigate the motion of the different PL-markers the total distance travelled for each marker was calculated by summing the Euclidean distance in 3-dimensional space between successive frames in each selected excerpt from the motion capture recordings. As shown in Figure 2, the bodily motion differed between emotions as well as between the emotional intensity displays, with greater distances travelled for exaggerated than normal expressions of happiness, sadness and anger. For all emotions and emotional displays, the hands were found to travel the greatest distance, which suggests that they were the most active body part.

Music. Audio recordings of solo piano pieces were used as musical material. Two pianists, one female and one male, both 27 years old, with 20 years of piano playing experience, were paid to perform four pieces each.[2] The pieces were selected according to the following criteria: (1) They should be short in duration, (2) relatively unknown (to musically untrained individuals) and (3) easy to interpret in different emotions (i.e., they should be relatively emotionally neutral and homogeneous in structure). One week before the recording, the pianists were informed that their task was to interpret the different pieces in a happy, content, sad and angry manner. During this week both of them practiced these performances. In the studio, pianists played on a Yamaha Clavinova CLP-150 and performances were recorded in MIDI format using that electronic piano. The recordings were done separately (i.e., pianists did not hear each others' performances). Each performance was repeated until the pianist judged it to be satisfactory (range 1–16 times) in the sense that the performance was error-free and produced with the best possible interpretation. The first 7s of each performance were chosen to be used as stimuli in a

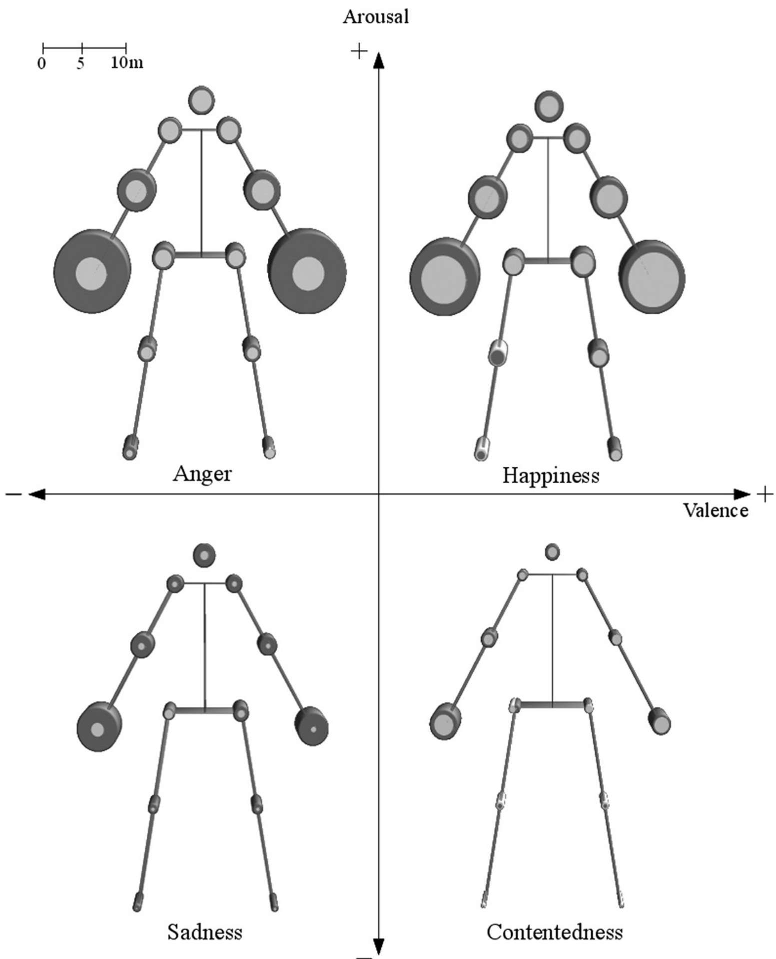

Figure 2. Total distance travelled for each PL-marker, averaged across actors both within and between dyads. Lightly shaded circles represent normal emotion displays; dark circles represent exaggerated emotion displays

pretest. These stimuli were normalized, using Adobe Audition, to a level of 24dB with an "Equal Loudness Contour," and logarithmic fade-in and fade-out were applied to the first and last 500ms of each excerpt.

Music and Video. A pretest was conducted to select auditory and visual stimuli to be combined in the main experiment. In this pretest, the 7s excerpts of all visual and auditory performances were presented in separate randomly ordered trials. The participants were asked to rate the emotions being expressed in these stimuli using a 2dES. The handling of this 2dES was practiced before the pretest in the same manner as for the main experiment (see below). After the pretest, the two videos and the three musical pieces with the highest score for recognition of each emotion were chosen. For the music, performances of Brahms (M1 and M2) and Schumann (M3) were selected for the experiment (see Figure 3).

Figure 3. Scores of the musical excerpts used in the experiment. a) First and second excerpt from Brahms' "Három Intermezzo", Opus 117 No. 1, Bars 1–3 (M1), and Bars 5–6½ (M2). b) Schumann's "*Grillen*", Fantasiestücke Opus 12 Bars 1–5 (M3)

Two-dimensional emotion space (2dES). Emotion evaluations were recorded using a 2dES (see Figure 4), which was implemented using Presentation software (Neurobehavioral Systems). Each quadrant of this 2dES was subdivided into 9 squares that could be selected via mouse click. The boxes at the point of origin indicate an almost neutral state, while boxes located at the margins indicate a higher or lower level of arousal and more positive or negative valence.

Procedure

The participants were given written instructions explaining the two concurrent tasks (decoy and experimental). For the decoy task, participants were required to decide whether a video, presented without music (V), was repeated within a group of five videos presented with music (VM). In the second task, which was the real experimental test, participants were asked to evaluate the emotion that they perceived in the interaction of the two persons depicted as PL-displays. Participants were instructed to concentrate on the videos alone and not to rate the music.

Before the experiment, a training phase was conducted to ensure that the participants could describe the relevant emotions accurately using the 2dES. Participants were initially presented

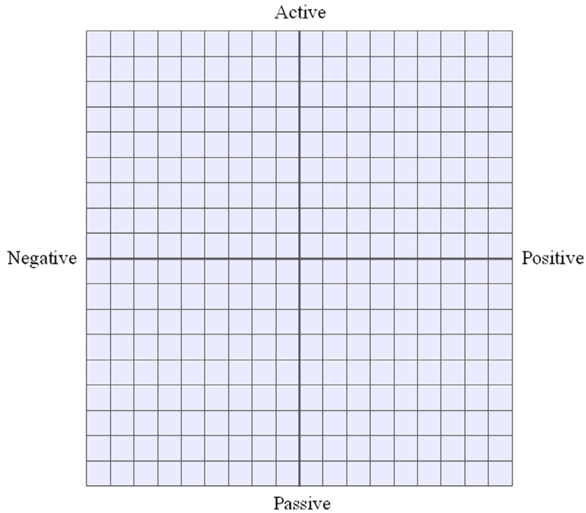

Figure 4. A schematic diagram of the two-dimensional emotion space (2dES) used for emotion evaluation. It was presented on the computer monitor in such a way that diagonal opposite corners were separated by 20 cm

with a paper version of this 2dES and were asked to explain its structural and conceptual organization using their own words. Subsequently they were asked to indicate specific emotional states (e.g., very happy or a bit sad) and to evaluate the emotions of characters in short sentences (unrelated to the experimental dialogue) using the paper version of the 2dES. The experimenter intervened only in the event of incorrect handling of the 2dES (e.g., the participant referred verbally to anger but indicated sadness by pointing with a finger). This training exercise also served to emphasize that the participants' emotion evaluations should be as precise as possible and that only one square could be selected. The training phase was concluded with two practice trials, consisting of two mini-blocks each, performed on the computer.

The two practice trials and the experiment itself were structured as follows. A visual cue ("!") was presented at the beginning of each mini-block for 1000ms. Within mini-blocks, a differently shaped visual cue ("+") was presented for 750ms before each stimulus. The first stimulus item, a PL-display presented without music (V), was followed by five displays presented with music (VM). The 2dES was displayed after each stimulus (V and VM) until the participant evaluated the perceived emotion by clicking in one square of the grid using the computer mouse. This mouse click triggered the presentation of the next stimulus after 750ms. At the end of each mini-block, text asking whether the first video was repeated among the five VM-stimuli was displayed on the monitor. Clicking on a "Yes" or "No" square with the computer mouse completed the mini-block.

Each participant encountered 32 mini-blocks during the experiment (i.e., 152 VM-stimuli; excluding the target-presentation VM-stimuli, see below). All mini-blocks and stimuli within them were presented in random order, with the following constraints: (1) no emotion was represented more than two times within the visual and auditory stimuli, (2) each musical piece (M1–M3) was used at most twice and (3) none of the five VM-stimuli were repeated within one mini-block. All VM-stimuli were shown twice within the experiment in addition to when they were used as target stimulus in the decoy task (i.e., in eight mini-blocks, the first video (V) was presented again in one VM-stimulus within the mini-block). These stimuli were presented once more in another mini-block as non-target stimuli (i.e., "non-target stimuli" for the decoy task). Only these "non-target" trials were used for further analyses, as the emotion evaluation of the target stimuli could be influenced by the emotional ratings of the V-stimulus.

The experiment was conducted in a dimly lit room, with videos viewed on a monitor (resolution: 1024 × 768) and audio heard over Sennheiser HD 270 headphones. Presentation software was used to control stimulus delivery and response collection. To ensure a reasonable duration of the experiment, only three of the four visual stimuli representing each emotion were presented to individual participants (i.e., as mentioned above, for each emotion the two best normal and exaggerated displays were selected to be used as test stimuli). Therefore, four subgroups of participants were formed, differing only in the visual stimuli that were used. Each participant was randomly assigned to one subgroup. Two rest breaks were included, one after the 10th and the other after the 21st block. Participants could terminate these breaks at will after 30s had elapsed.

To lessen the likelihood that the participants would form the belief that the study focused on music's impact on visual emotion perception, the entire testing procedure was conducted in the laboratories of the Psychology Department at the Max Planck Institute for Human Cognitive and Brain Sciences (which are located in a different building to the laboratories of the music group). After the experiment, participants were given a questionnaire to assess the familiarity of the musical pieces and to evaluate the tasks constituting the experiment itself. The participants were also asked to evaluate their performances and to guess what the research question might be.

Results

Emotion evaluations of the PL-displays were analysed by examining the quadrants of the 2dES chosen by the participants.[3] The selection of the correct quadrant (e.g., the upper right quadrant for a happy interaction) was coded as 1 while the selection of any other quadrant was coded as 0. As these emotion evaluation data were proportional in nature, an arcsine-transformation was performed on them after each participant's mean value in each condition was calculated. A repeated-measures ANOVA with Emotion (happiness, anger, contentedness and sadness), Display (normal and exaggerated), Music (emotionally compatible and incompatible with the visual stimuli) and Piece (the different musical pieces performed by the pianists) as the within-subjects factors was conducted on the emotion evaluation accuracy data. The criterion for statistical significance was set at $\alpha = .05$, with the Greenhouse-Geisser correction applied when the degrees of freedom numerator exceeded one. Neither the main effect of Piece nor its interaction with the other factors was statistically significant. Furthermore, inspection of post-test questionnaires revealed that none of the participants reported familiarity with the musical pieces. The factor of Piece will therefore not be discussed further.

A significant main effect was found for Emotion ($F_{3,45} = 10.322$, $p < .001$, $\eta^2 = .408$) as well as significant two- and three-way interactions (see below). As shown in Figure 5, all emotions were evaluated correctly above chance with the highest accuracy found for anger, followed by happiness, sadness and contentedness. Pairwise comparisons ($\alpha = .05$, Bonferroni corrected) showed significant differences in emotion evaluation accuracy between anger and contentedness ($p < .001$) as well as between anger and sadness ($p < .001$). No other significant differences were found.

A significant Emotion × Music interaction ($F_{3,45} = 6.781$, $p < .01$, $\eta^2 = .311$) indicates that some emotions were evaluated more correctly when the videos were presented with

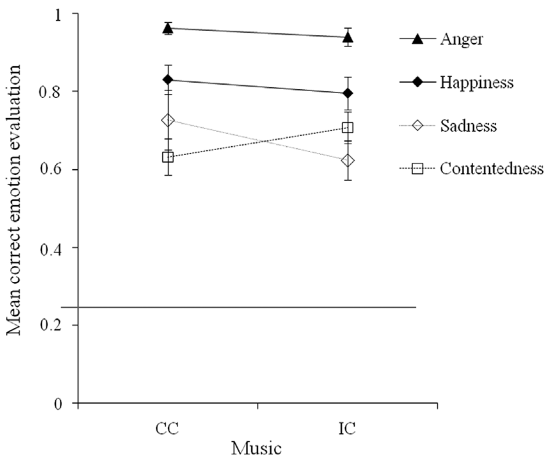

Figure 5. Mean correct emotion evaluations of happiness, contentedness, sadness and anger in PL-displays with emotionally compatible (CC) or incompatible (IC) music. Error bars indicate standard error and the horizontal line represents chance level performance (.25%)

emotionally compatible than incompatible music. This was confirmed by conducting simple main effect analyses of the effect of Music for each emotion separately. As can be seen in Figure 5, portrayals of happiness and sadness were correctly classified significantly more often when presented with compatible than incompatible music ($F_{1,15} = 4.601$, $p < .05$, $\eta^2 = .235$ for happiness and $F_{1,15} = 7.046$, $p < .05$, $\eta^2 = .320$ for sadness). These results are in line with our hypothesis that emotionally compatible music increases evaluation accuracy for emotional PL-stimuli. Emotional displays of contentedness on the other hand were evaluated significantly less correctly when presented with emotionally compatible music than with incompatible music ($F_{1,15} = 6.153$, $p < .05$, $\eta^2 = .291$). For the expressions of anger no significant difference in classification accuracy were found for compatible and incompatible music.

The interaction of Emotion × Display was also found to be significant ($F_{3,45} = 13.278$, $p < .001$, $\eta^2 = .470$). Simple main-effect analyses of the effect of Display revealed a significant impact of this factor on emotion evaluations of happiness, contentedness and sadness. As shown in Figure 6, happiness and sadness were correctly evaluated significantly more often when portrayed in an exaggerated way than in a normal manner ($F_{1,15} = 5.827$, $p < .05$, $\eta^2 = .280$ for happiness and $F_{1,15} = 14.603$, $p < .01$, $\eta^2 = .493$ for sadness). For the expression of contentedness, the exaggerated display decreased evaluation accuracy significantly compared to the normal emotion display ($F_{1,15} = 20.021$, $p < 001$, $\eta^2 = .572$). For the expression of anger no significant impact of Display was observed ($F_{1,15} = .769$, p =.394, $\eta^2 = .049$).

It was also hypothesized that exaggerated emotion displays will be less affected by music than the normal ones. The significant three-way interaction of Emotion × Display × Music ($F_{3,45} = 2.816$, $p < .05$, $\eta^2 = .158$) indicates such an effect for some emotions. To break down this interaction, simple main-effect analyses were conducted for the interaction of Music × Display for each emotion separately (see Figure 6). A significant effect was found only for sadness, with the impact of music (i.e., greater evaluation accuracy for the visual stimuli with compatible than incompatible music) being significantly higher for normal displays compared to exaggerated ones ($F_{1,15} = 4.934$, $p < .05$, $\eta^2 = .247$).

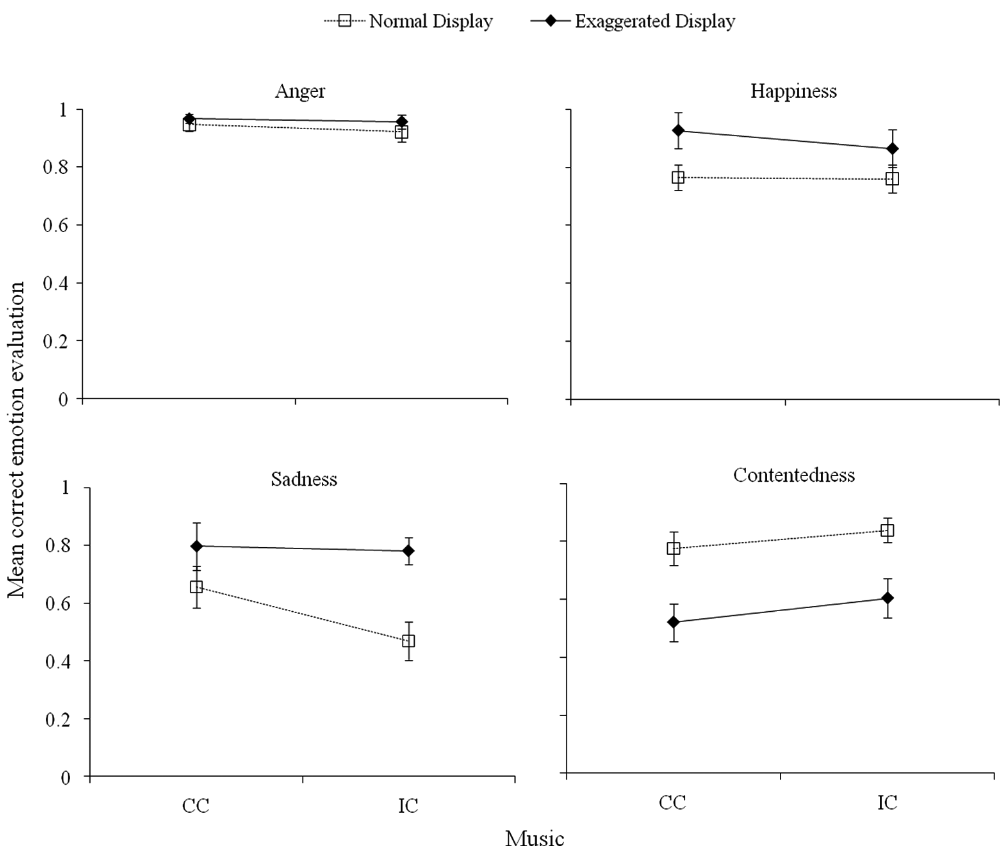

Figure 6. Mean correct emotion evaluations of normal and exaggerated expressions of happiness, contentedness, sadness and anger in PL-displays with emotionally compatible (CC) and incompatible (IC) music. Error bars indicate standard error

An additional analysis was conducted to check whether the above effects were modulated by the beliefs of the participants about the research question. The post-test questionnaire revealed that seven participants believed that the study examined the visual perception and/or recognition of emotions in PL-stimuli, indicating that they associated music as a source of interference. The remaining nine participants correctly suspected that the experiment investigated the impact of music on visual perception.

A mixed-ANOVA with Emotion, Display, Music and Piece as within-subjects factors and Belief about the research question as a between-subjects factor yielded a significant Emotion × Display × Music × Belief interaction ($F_{3,42} = 4.858$, $p < .01$, $\eta^2 = .255$). This interaction was further examined by conducting simple main effect analyses for the interaction of Music × Belief for every combination of Emotion × Display. Significant effects were found for normal expressions of sadness ($F_{1,14} = 5.064$, $p < .05$, $\eta^2 = .266$) and exaggerated expressions of contentedness ($F_{1,14} = 4.930$, $p < .05$, $\eta^2 = .260$), which were the two least accurately evaluated emotion expressions in this study. For normal expressions of sadness, the effect of music, which increased evaluation accuracy when it was emotionally compatible with the video, was significantly higher for participants who did not report awareness of the research question than for

Table 1. Mean correct emotion evaluation (in percentage) of normal (N) and exaggerated (E) emotion displays presented with emotionally compatible (CC) and incompatible (IC) music, shown separately for participants who did not report awareness of the research question (Unaware) and participants who did (Aware). Standard error is shown in parentheses

Stimulus display	Stimulus evaluation			
	Unaware (n = 7)		Aware (n = 9)	
	CC	IC	CC	IC
Happiness				
N	74 (.059)	70 (.072)	79 (.067)	81 (.062)
E	100 (.000)	94 (.047)	87 (.109)	81 (.108)
Contentedness				
N	63 (.087)	75 (.066)	89 (.048)	91 (.047)
E	48 (.110)	43 (.100)	56 (.082)	74 (.061)
Sadness				
N	76 (.078)	44 (.087)	57 (.110)	49 (.101)
E	85 (.092)	81 (.065)	76 (.130)	76 (.070)
Anger				
N	94 (.030)	92 (.048)	95 (.037)	93 (.049)
E	98 (.024)	98 (.024)	96 (.025)	94 (.039)

those who did. The "unaware" participants also evaluated exaggerated expressions of contentedness more accurately when presented with emotionally compatible than incompatible music. For the other "aware" participants the opposite trend was observed.

Taken together, these findings suggest that our main results are not attributable to response bias engendered by awareness of the research question, since the performance of "unaware" participants conformed more closely to our hypotheses than the performance of "aware" participants.

Discussion

The present study was designed to investigate the impact of music on the perception of emotion in kinematic displays of interpersonal interaction. Concerning the bodily emotion expressions themselves, all were evaluated correctly at levels above chance. This corroborates the findings of Clarke et al. (2005), showing that observers are able to perceive emotions correctly in kinematic displays of interpersonal dialogue. The results are also in line with other studies (e.g., Wallbott, 1998), demonstrating that the body is a good channel for communicating the emotions of happiness, sadness and anger in an effective way. The same seems to apply for contentedness, although fewer other studies are available for comparison to strengthen this conclusion.

Atkinson et al. (2004) demonstrated that exaggerated displays by individual actors can increase the accuracy of emotion evaluation. The results of the present study indicate that this effect can also be found for bodily expressions of certain emotions during social interaction. As for the findings of Atkinson et al. (2004), happiness was perceived more accurately when expressed in an exaggerated manner. Contrary to the findings of Atkinson et al. (2004), an increase in recognition performance was also found for exaggerated displays of sadness in the current study. Atkinson et al. (2004) attributed the absence of such an effect in their study to the actors' tendency to express exaggerated portrayals of all emotions, including sadness, with faster and larger motions. Calculating the total distance travelled for each marker of the PL-stimuli presented in our study revealed that our actors indeed increased the level of activity when they expressed exaggerated happiness, anger and sadness (see Figure 2). The increase in recognition performance for exaggerated expressions of sadness found in our study may therefore be due to the social setting of our PL-stimuli. Clarke et al. (2005) reported sadness to be an essentially social emotion, as were love, joy and fear, as the evaluation accuracy of these emotions was found to be affected by the presence of a partner (see also Rose & Clarke, 2009). Although the presence of an interacting partner did not affect the perception of *normal* expressions of sadness in the study of Clarke et al. (2005), it may have increased the evaluation accuracy of the *exaggerated* expressions of this emotion in the present study. Even though further investigation is necessary to support this interpretation, the differences found in the current study reinforce the claim that it is beneficial to investigate emotional expressions in social settings.

The absence of a significant gain in evaluation accuracy for exaggerated displays of anger in the current study could be due to the very high evaluation accuracy found for normal expressions of this emotion (i.e., a ceiling effect). In the study of Atkinson et al. (2004), normal expressions of anger were evaluated less accurately compared to the other emotions under investigation (i.e., happiness, sadness, fear and disgust). In the current study, however, portrayals of anger were evaluated most accurately, regardless of whether they were performed in a normal or exaggerated way. This finding may be attributable to the social setting provided by our dyadic interaction displays, in which the observer can see the emotional reaction of each actor to the other's emotional action. The results of Clarke et al. (2005) provide evidence for such a boosting effect of social interaction for some emotions. Although anger was recognized best in their study, as in ours, its perception was not found to be affected by the social context (i.e., an interacting couple, a single actor from this couple, or a single actor plus his or her reflection). Note that this result does not necessarily contradict the notion of boosting effects of social interaction. All stimuli in the study by Clarke et al. (2005) were created in the context of a social interaction, and therefore emotional reactions were presumably visible in the motions of each individual person even when their co-actor was invisible.

It was hypothesized that the evaluation accuracy of bodily emotion expressions should be increased when presented with emotionally compatible music. Results indicated that music affected the perception of dyadic interactions expressing happiness, sadness and contentedness, even though participants were instructed to evaluate the emotion of the visual stimuli alone. In accordance with our hypothesis, happiness and sadness were evaluated more accurately when presented with emotionally compatible music (i.e., happy interactions combined with happy music and sad interactions combined with sad music). Unexpectedly, portrayals of contentedness were evaluated less accurately when presented with compatible music. This decrease in evaluation accuracy could be due to ambiguity in the emotional character of the music itself, as musical performances of contentedness were often perceived as expressions of sadness (see pretest ratings in the Appendix, Table A2). Such emotion confusion is consistent with Juslin's (2001) observation that musical performances of these two emotions are frequently found to be mistaken for one another. Given this frequent confusion of contentedness and sadness, the current finding that contentedness was evaluated less accurately when presented with music that was intended to be emotionally compatible than with incompatible (angry) music, may reflect a stronger impact of music on visual stimuli when both differ on only one dimension (since sadness and contentedness differ only in valence).

Our second hypothesis was motivated by Cohen's (1993) finding that only ambiguous stimuli (short FL-movies) were found to be affected by music. To investigate whether such an effect can also be found for PL-stimuli of social nonverbal communication, normal and exaggerated emotion portrayals were used in the current study. As mentioned above, our findings confirm the observation by Atkinson et al. (2004) that exaggerated emotion portrayals can increase the evaluation accuracy of emotion expression, and thus reduce the ambiguity of such visual stimuli. An increase in recognition performance due to exaggerated emotion expression was found for happiness and sadness, with a major gain for sadness. The perception of happiness and sadness should therefore be affected less by the music when displayed in an exaggerated manner, compared to normal emotion expression. Such an effect was found for sadness. While normal emotion portrayals of sadness were evaluated significantly less correctly when presented with emotionally incompatible music, the evaluation ratings for the exaggerated displays did not differ reliably between the two music conditions. The absence of such an effect for happiness can be explained by the smaller gain in evaluation accuracy for exaggerated expressions of happiness relative to sadness. Additional results indicated that music's impact on the visual perception of a kinematic display can be affected by the emotional clarity of the display. Notably, music was not found to affect the perception of anger. Portrayals of this emotion can be considered to be less ambiguous as they were evaluated most accurately relative to expressions of happiness, sadness and contentedness. These results for sadness and anger indicate that the restricted impact of music found by Cohen (1993) for emotionally ambiguous FL-videos generalizes to kinematic displays of some emotions.

In summary, the results of the present study demonstrate that the perception of visual nonverbal emotion communication in PL-displays can be affected by music. These effects of music are generally predicated on the ability of the performer to communicate emotions by varying parameters such as intensity and tempo (see Juslin & Timmers, 2010). We observed that kinematic displays of happiness and sadness were evaluated more accurately when accompanied by emotionally compatible music, while contentedness was discerned best when presented with emotionally incompatible music. Anger was found to be resistant to the effects of music. A reduced impact of music on visual emotion perception was also found for exaggerated expressions of sadness. Happiness and sadness were evaluated more accurately when expressed in an exaggerated manner, whereas contentedness was evaluated less accurately when exaggerated. For the exaggerated, less ambiguous expressions of sadness, a reduced impact of music was found compared to normal expressions of this emotion. Our analyses of participants' beliefs suggested that the above effects were not due to response biases related to awareness of the research question.

Our results support and extend previous findings, as the usage of PL-stimuli allows the observed impact of music to be interpreted as a cross-modal interaction of auditory information, specifically with the kinematics of nonverbal communication. Furthermore, it was demonstrated that these effects are constrained by the degree of emotional ambiguity in the visual displays. It would be profitable for future studies to concentrate on further variables in the attempt to understand the cross-modal interaction of music and visual perception in greater depth. A promising starting point could be the investigation of the dimensions of valence and arousal separately from one another, as our results indicate a higher impact of music on visual emotion perception when both sources differ only in one dimension.

Acknowledgements

The authors wish to thank Jan Bergmann, Andreas Weber, Cornelia Schulze and Nathanaël Lécaudé for technical assistance. The study was part of author RK's diploma thesis, conducted under the supervision of author PEK, and was supported by the Max Planck Society. Author RK was supported by a stipend from the MaxNetAging International Research Network while working on the manuscript.

Notes

1. One participant reported mild attenuation of high frequencies in one ear, but this is unlikely to have affected his ability to perform the experimental task.
2. The four musical pieces performed by the two pianists (A and B) were:
   (A, B) Felix Mendelssohn Bartholdy "Lied ohne Worte", Opus 67 No. 3 (Bars 1–11½)
   (A, B) Ludwig van Beethoven "Bagatelle", Opus 126 No. 5 (Bars 1–17)
   (A) Franz Peter Schubert "Impromptu", Opus 90 D 899 No. 2 (Bars 1–24)
   (A) Johann Sebastian Bach "Toccata", Partita No. 6 BWV 830 (Bars 1–15)
   (B) Robert Schumann "Grillen", Fantasiestücke Opus 12 (Bars 1–17)
   (B) Johannes Brahms "Három Intermezzo", Opus 117 No. 1 (Bars 1–8½).
   Note that each pianist was allowed to suggest pieces on their own following the above mentioned criteria.
3. The squares within the quadrants were used to investigate perceived emotional intensity. Analyses examining the squares selected within the quadrants of the 2dES indicated that exaggerated emotion expressions were evaluated to be more emotionally intense (i.e., squares nearer the periphery of the 2dES were selected) relative to normal expressions. In the interest of conciseness we do not discuss these results further in this article.

References

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, *33*(6), 717–746.

Atkinson, A. P., Tunstall, M. L., & Dittrich, W. H. (2007). Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. *Cognition*, *104*, 59–72.

Bolivar, V. J., Cohen, A. J., & Fentress, J. C. (1994). Semantic and formal congruency in music and motion pictures: Effects on the interpretation of visual action. *Psychomusicology*, *13*, 28–59.

Chouchourelou, A., Matsuka, T., Harber, K., & Shiffrar, M. (2006). The visual analysis of emotional actions. *Social Neuroscience*, *1*(1), 63–74.

Clarke, E. F. (2004). Empirical methods in the study of performance. In E. F. Clarke & N. Cook (Eds.), *Empirical musicology: Aims, methods, prospects* (pp. 77–102). New York: Oxford University Press.

Clarke, T. J., Bradshaw, M. F., Field, D. T., Hampson, S. E., & Rose, D. (2005). The perception of emotion from body movement in point-light displays of interpersonal dialogue. *Perception*, *34*(10), 1171–1180.

Cohen, A. J. (1993). Associationism and musical soundtrack phenomena. *Contemporary Music Review*, *9*(1), 163–178.

Cohen, A. J. (2005). How music influences the interpretation of film and video: Approaches from experimental psychology. In R. Kendall & R. Savage (Eds.), *Selected reports in ethnomusicology: Perspectives in systematic musicology* (Vol.12, pp. 15–36). Los Angeles, CA: Ethnomusicology Publications, UCLA.

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musician's body movements. *Music Perception*, *24*(5), 433–454.

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, *21*, 103–113.

Davidson, J. W., & Correia, J. S. (2002). Body movement. In R. Parncutt, & G. E. McPherson (Eds.), *The science and psychology of music performance. Creative strategies for teaching and learning* (pp. 237–250). Oxford: Oxford University Press.

Dittrich, W. H. (1993). Action categories and the perception of biological motion. *Perception*, *22*(1), 15–22.

Dittrich, W. H., Troscianko, T., Lea, S. E. G., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, *25*(6), 727–738.

Dissanayake, E. (2006). Ritual and ritualization: Musical means of conveying and shaping emotion in humans and other animals. In S. Brown & U. Volgsten (Eds.), *Music and manipulation* (pp. 31–56). Oxford and New York: Berghahn Books.

Ekman, P. (1993). Facial expression and emotion. *American Psychological Association*, *48*(4), 376–379.

Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion* (pp. 45–60). Sussex: John Wiley.

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*, 68–91.

Gabrielsson, A., & Juslin, P. N. (2003). Emotional expression in music. In R. J. Davidson, K. R. Scherer, & H. H Goldsmith (Eds.), *Handbook of affective sciences* (pp. 503–534). New York and Oxford: Oxford University Press.

Gabrielsson, A., & Lindström, E. (2001). The influence of musical structure on emotional expression. In P. N. Juslin and J. A. Sloboda (Eds.), *Music and Emotion: Theory and research* (pp. 223–248). Oxford: Oxford University Press.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309–337). Oxford: Oxford University Press.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814.

Juslin, P. N., & Timmers, R. (2010). Expression and communication of emotion in music performance. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, and applications* (pp. 453–489). Oxford: Oxford University Press.

Lipscomb, S. D., & Tolchinsky, D. E. (2005). The role of music communication in cinema. In D. Miell, R. MacDonald, & D. Hargreaves (Eds.), *Musical Communication* (pp. 383–404). Oxford: Oxford University Press.

Livingstone, S. R., Thompson, W. F., & Russo, F. A. (2009). Facial expressions and emotional singing: A study of perception and production with motion capture and electro-myography. *Music Perception*, *26*(5), 475–488.

Madison, G. (2000). Properties of expressive variability patterns in music performances. *Journal of New Music Research*, *29*(4), 335–356.

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, *82*(2), B51–B61.

Rose, D., & Clarke, T. J. (2009). Look who's talking: Visual detection of speech from whole-body biological motion cues during emotive interpersonal conversation. *Perception*, *38*, 153–156.

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General*, *112*(4), 585–615.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, *57*(3), 493–502.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *115*(1), 102–141.

Schubert, E. (1996). Continuous response to music using a two-dimensional emotion space. In Proceedings of the 4th International Conference of Music Perception and Cognition (pp. 263–268). Montreal, Canada: ICMPC.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, *21*(4), 561–585.

Sloboda, J. A., & Juslin, P. N. (2001). Psychological perspectives on music and emotion. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 71–104). Oxford: Oxford University Press.

Thompson, W. F., Russo, F. A., & Quinto, L. (2008). Audio-visual integration of emotional cues in song. *Cognition & Emotion*, *22*(8), 1457–1470.

Troje, N. F., Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: Effects of structural and kinematic cues. *Perception & Psychophysics*, *67*(4), 667–675.

Van den Stock, J., Peretz, I., Grèzes, J., & De Gelder, B. (2009). Instrumental music influences recognition of emotional body language. *Brain Topography*, *21*(3–4), 216–220.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, *101*, 80–113.
Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, *28*(6), 879–896.
Zehnder, S. M., & Lipscomb, S. D. (2006). The role of music in video games. In P. Vorderer & J. Bryant (Eds.), *Playing videos games: Motives, responses, and consequences* (pp. 241–257). Mahwah, NJ: Lawrence Erlbaum Associates.

Appendix

Table A1. Dialogue between actor A and B used for creation of the visual PL-stimuli (left), and a translation of the German text into English (right).

Actor	Text	
A	Es ist ein langer Weg nach Russland.	It is a long way to Russia.
B	Ich hab nicht gedacht, dass es so weit wär'.	I did not believe that it would be so far.
A	Aber es gibt einen Ausflug nach Moskau.	But there is a tour to Moscow.
B	Davon hat mir aber niemand erzählt.	Nobody told me about that.
A	Ja, die Botschaft hat alles dafür organisiert.	Yes, the embassy has organized everything.
B	Ich dachte das hätten unsere Kollegen getan.	I thought our colleagues had done that.
A	Wenn der Zug dort extra halten kann?	Perhaps the train could make an extra stop there?
B	So etwas könnte organisiert werden.	Something like that could be organized.
A	Genau!	Right!

Table A2. Pre-test ratings: Evaluations (in percentage) of the first and second theme of the Brahms performances (M1 and M2) and the Schumann performances (M3) expressing happiness (H), contentedness (C), sadness (S) and anger (A). Bold figures indicate correct emotion evaluation.

Stimulus	Stimulus evaluation			
	H	C	S	A
M1				
H	**75**	17	0	8
C	17	**42**	42	0
S	8	33	**50**	8
A	50	17	8	**25**
M2				
H	**75**	0	17	8
C	17	**58**	17	8
S	8	42	**25**	25
A	58	8	0	**33**
M3				
H	**42**	0	0	58
C	25	**8**	33	33
S	8	0	**83**	8
A	25	0	8	**67**