

# OH HAPPY DANCE: EMOTION RECOGNITION IN DANCE MOVEMENTS

Birgitta Burger, Marc R. Thompson, Suvi Saarikallio, Geoff Luck,  
Petri Toiviainen

Finnish Centre of Excellence in Interdisciplinary Music Research, University of Jyväskylä, Finland  
birgitta.burger@jyu.fi

## Abstract

Movements are capable of conveying emotions, as shown for instance in studies on both non-verbal gestures and music-specific movements performed by instrumentalists or professional dancers. Since dancing/moving to music is a common human activity, this study aims at investigating whether quasi-spontaneous music-induced movements of non-professional dancers can convey emotional qualities as well. From a movement data pool of 60 individuals dancing to 30 musical stimuli, the performances of four dancers that moved most notably, and four stimuli representing happiness, anger, sadness, and tenderness were chosen to create a set of stimuli containing the four audio excerpts, 16 video excerpts (without audio), and 64 audio-video excerpts (16 congruent music-movement combination and 48 incongruent combinations). Subsequently, 80 participants were asked to rate the emotional content perceived in the excerpts according to happiness, anger, sadness, and tenderness. The results showed that target emotions could be perceived in all conditions, although systematic mismatches occurred, especially with examples related to tenderness. The audio-only condition was most effective in conveying emotions, followed by the audio-video condition. Furthermore in the audio-video condition, the auditory modality dominated the visual modality, though the two modalities appeared additive and self-similar.

**Keywords:** music-induced movement, emotion, perception

## 1. Introduction

On a daily basis, humans use body movements as an important means of nonverbal communication. Body postures and movements can convey different kinds of information, for instance related to the mental or physical state or personality traits, or to accompany and emphasize speech. It has been argued that speech and movement/gestures are tightly connected and co-occur, as they underlie the same cognitive processes (Iverson & Thelen, 1999; McNeill, 1992). Furthermore, body movements can convey emotions. Various studies have investigated the capability of movement to express emotions and have shown that distinct features of human movement are related to emotion categories. De Meijer (1989), for example, asked observers to attribute emotional charac-

teristics to movements of actors who performed several movement patterns differing in general features, such as trunk or arm movement, velocity, and spatial direction and found that different emotion categories were associated with different movement characteristics. Wallbott (1998) conducted a study in which he used a scenario-based approach with professional actors performing certain emotions. He found movement features characteristic for different emotion categories and computed a discriminant analysis that could, significantly above chance level, classify the emotions correctly. Atkinson, Dittrich, Gemmell, and Young (2004) compared static vs. dynamic whole body expressions and full-light vs. point-light displays and found that all of them could

communicate emotions, though the recognition rate and misclassification patterns differed for individual emotions. Pollick, Paterson, Bruderlin, and Sanford (2001) investigated visual perception of emotions in simple arm movements shown as point-light displays, such as drinking and knocking. They found that arm movements could communicate emotions, though observers tended to confuse similar emotions. Gross, Crane, and Fredrickson (2010) studied the perception of knocking movements performed by actors who were subjected to an emotion induction task. The results indicated a limited recognition rate, especially for positive emotions, and systematic confusions between several emotions. Besides in acted emotion approaches, emotion recognition has been studied in other contexts as well, such as in gait. Montepare, Goldstein, and Clausen (1987), for instance, showed that happiness, anger, and sadness could be successfully recognized in walking patterns. Research in linguistics has investigated integration of auditory and visual information in emotion perception using face-voice stimuli. Such studies showed that usually the visual information dominates (Collignon et al., 2008), and that bimodal stimuli can be integrated even if they display incongruent combinations of emotions (De Gelder & Vroomen, 2000; Massaro & Egan, 1996).

Emotions are an essential component of musical expression (e.g., Gabrielsson & Lindström, 2010) and have been investigated in a large number of music-related studies. According to Krumhansl (2002), people report that their primary motivation for listening to music is its emotional impact. Various rating experiments have shown that listeners are able to perceive emotional content in music in a consistent fashion (e.g., Balkwill & Thompson, 1999; Eerola & Vuoskoski, 2011; Gabrielsson & Juslin, 1996; Schubert, 1999; Zentner, Grandjean, & Scherer, 2008).

Musical emotions are conveyed not only by the music itself, but also through movement. While movements are required, for example, to produce sounds when playing a musical instrument, studies have shown that there are certain additional movements that are not used for the actual sound production, but for

conveying emotions and expressivity (e.g., Wanderley, Vines, Middleton, McKay, & Hatch, 2005). Davidson (1993) conducted a study in which observers rated expressive movements of violinists and pianists. Results indicate that visual information more clearly communicated the expressive manner of the musician than sound alone and sound and vision presented together. Vines, Krumhansl, Wanderley, and Levitin (2006) examined clarinetists' abilities of communicating tension to observers and found that auditory and visual signals evoked different perceptions of tension, with the sound dominating the judgments. Additionally, their results indicated that the audiovisual presentation increased the perceived tension compared to the audio-only and video-only conditions, suggesting that participants integrated both signals. Dahl and Friberg (2007) studied marimba, saxophone, and bassoon players performing with different emotional intentions and presented observers with only the visual elements of the performances. Observers could detect the happy, angry, and sad performances successfully, but failed with the fearful ones. Sörgjerd (2000) investigated a clarinetist and a violinist who performed a piece of music with different emotional intentions, and reported that happiness, anger, sadness, and fear were better identified than tenderness and solemnity. However, no significant differences for the presentation condition (audio-only, movement-only, audio+movement) were found. Petrini, McAleer, and Pollick (2010) found that in audiovisual presentations of musicians playing with different emotional characteristics the sound dominated the visual signal, both in emotionally matching and mismatching stimuli. Furthermore, when participants were asked to focus on the visual information of the audiovisual stimuli, their ability to correctly identify the emotion decreased in case of emotionally incongruent stimuli, whereas it was unaffected by the video information, when participants were asked to focus on the audio information.

More direct links between music and emotion-specific movement have been investigated in research on dance, in which movement is the only way to convey expressivity and emotion. Several studies showed that dance

movement could successfully communicate emotions to observers, both in regular video and in stick-figure animations (Boone & Cunningham, 1998; Dittrich, Troscianko, Lea, & Morgan, 1996; Lagerlöf & Djerf, 2009; Walk & Homan, 1984). Common to these studies was that they used professional dancers (or actors) who were explicitly asked to express the emotions while dancing.

Listening to music make people move spontaneously, for example by rhythmically synchronizing with the pulse of the music by tapping the foot, nodding the head, moving the whole body in various manners, or mimicking instrumentalists' gestures (Leman & Godøy, 2010; Leman, 2007). Studies investigating music-induced movement have suggested such movements to be related to personality (Luck, Saarikallio, Burger, Thompson, & Toiviainen, 2010), mood (Saarikallio, Luck, Burger, Thompson, & Toiviainen, 2013), or musical features (Burger, Thompson, Saarikallio, Luck, & Toiviainen, 2013; van Dyck et al., 2013). In a recent study (Burger, Saarikallio, Luck, Thompson, & Toiviainen, 2013), we could also establish links between movements and the emotional content of the music to which the participants were dancing. In that experiment, we asked 60 participants to move to different pop music stimuli and recorded their movements with an optical motion capture system. The computational analysis of the movement data coupled with a perceptual evaluation of the emotions expressed in the music revealed characteristic movement features for the emotions happiness, anger, sadness, and tenderness.

While there appears to be correlations between perceived emotions in music and movement characteristics, it is neither clear whether music-induced movement can convey emotional content to observers and which emotions can be communicated, nor how auditory and visual information would interact in this process. Therefore, we designed a perceptual experiment using a subset of the movement data collected in the previous experiment and asked observers to rate various stick figure clips regarding the emotions conveyed by these clips. We restricted this study to the use of perceived emotions as opposed to felt

emotions, since previous literature has mostly focused on the former (about the importance to distinguish between perceived and felt emotions, see Evans & Schubert, 2008; Gabrielsson, 2002), and we used perceived emotions in the previous study. We decided to include three presentation conditions: audio-only, video-only, and audio and video combined, as was done in studies about musician's movements (Davidson, 1993; Petrini et al., 2010; Sörgerd, 2000; Vines et al., 2006) and in research on face-voice stimuli (Collignon et al., 2008; De Gelder & Vroomen, 2000; Massaro & Egan, 1996). Besides including the correct combinations of music and movement (i.e., the movements actually performed to that music during the previous experiment), we wanted to further examine the audiovisual integration and generated a set of incongruent stimuli (i.e., combining the movements with another song used in the experiment expressing a different emotion), as was done in Petrini et al. (2010), for instance. Previous studies suggest different scenarios regarding the influence of music and movement on the judgments. However, in line with the results by Petrini et al. (2010) and Vines et al. (2006), we hypothesize that the audio-video condition will receive higher recognition rates than the two unimodal conditions (at least for the congruent stimuli), and audio will dominate the perception of both the congruent and the incongruent audio-video examples.

## 2. Method

### 2.1. Participants

Eighty university students, aged 19-36, participated in this study (53 females, average age: 24.7, SD of age: 3.4). Fifty-two participants took some kind of dance lessons and 75 participants reported to like movement-related activities, such as dance and sports. Thirty-seven participants reported to go out dancing more than once a month.

### 2.2. Stimuli

The stimuli used in this experiment were selected from a motion capture data pool of 60

participants dancing to 30 different musical excerpts that was collected in a previous study (Burger, Saarikallio, et al., 2013). Based on results of a rating experiment conducted within this previous study, four musical excerpts were determined that most clearly conveyed one of the following (target) emotions happiness, anger, sadness, and tenderness. Subsequently, two female and two male participants (from now referred to as “dancers”), were chosen based on their movement characteristics (i.e., participants whose performances received highest values regarding several movement features, such as speed, acceleration, area covered, or complexity). This yielded 16 (4x4) combinations of stick figure dance performances.

The four musical stimuli were of different tempi (100 bpm, 105 bpm, 113 bpm, 121 bpm), so we adjusted their tempi to 113 bpm by time-stretching three of them using the Audacity software<sup>1</sup>, to eliminate the effect of tempo on the perceived emotions. Likewise, the movement data was resampled by the appropriate ratio using the Matlab Mocal Toolbox (Toiviainen & Burger, 2013). Besides this, the movement data were rotated to be visible from the front with respect to the average locations of the hip markers (for more information on the marker locations see Burger et al., 2013). In all videos, the stick figures were plotted in black on white background. QuickTime Player 7 (Pro version) was used to combine the time-shifted audio and video material in all possible combinations yielding 16 congruent stimuli and 48 incongruent stimuli. It was checked that the combination appeared synchronized. All stimuli used in the experiment were trimmed to 20 seconds.

### 2.3. Apparatus

To gather perceptual ratings, a special patch was created in Max/MSP 5, a graphical programming environment, running on Max OS X. The patch used QuickTime Player 7 to play back the video material. The setup enabled the participants to repeat excerpts as often as they wished, to move forward at their own speed, and to take breaks at any moment of the ex-

periment. The stimuli were played back through studio quality headphones (AKG K141 Studio). The participants could themselves adjust the volume to a preferred level.

### 2.4. Procedure

In the beginning of the experiment, a short questionnaire was filled in to gather information about participants' gender, age, dance training, and movement and dance activities. The experiment was divided into three sections: one section containing the four (time-stretched) audio clips, a second section containing the 16 silent (and time-shifted) video clips (four dancers moving to the four different musical stimuli), and a third section containing the 64 (time-shifted) audio-video clips (16 congruent and 48 incongruent combinations). To avoid any effect of order, the three sections were presented in random order to the participants. Within each section, the clips were randomized as well. Participants were accomplishing the experiment individually. They were instructed to rate the emotions expressed in the clips (perceived emotions) on seven-step scales for Happiness, Anger, Sadness, and Tenderness. Preceding each section, there was a practice part of one example to get familiar with the interface, the type of stimuli, and the rating scales used. In the beginning of the experiment, participants were explicitly told to rate according to the emotions expressed in the clips (as opposed to felt emotions). The participants were also advised to take breaks in between if they felt like it. The total duration of the experiment was between 45 and 90 minutes. After completing the experiment participants were rewarded with a movie ticket.

## 3. Results

Outlier detection was performed as the first step of the analysis by calculating the mean inter-subject correlation for each participant (taking all ratings of the three conditions) and each rating scale (happiness, anger, sadness, and tenderness) separately. This yielded a measure of how similarly the participants were rating in relation to each other on each scale.

---

<sup>1</sup> <http://www.audacity.sourceforge.net>

We obtained overall positive correlations, apart from one participant who correlated negatively on two scales (happiness and tenderness) with the other participants. We therefore decided to eliminate this participant from further analysis.

Next, we checked for rating consistency between participants by calculating intraclass correlations (cf., Shrout & Fleiss, 1979) for each rating scale separately. All six correlations coefficients were highly significant (between  $r = .95$  and  $r = .97$ ,  $p < .001$ ), suggesting that participants' ratings for each scale were similar enough to average across them for any further analysis.

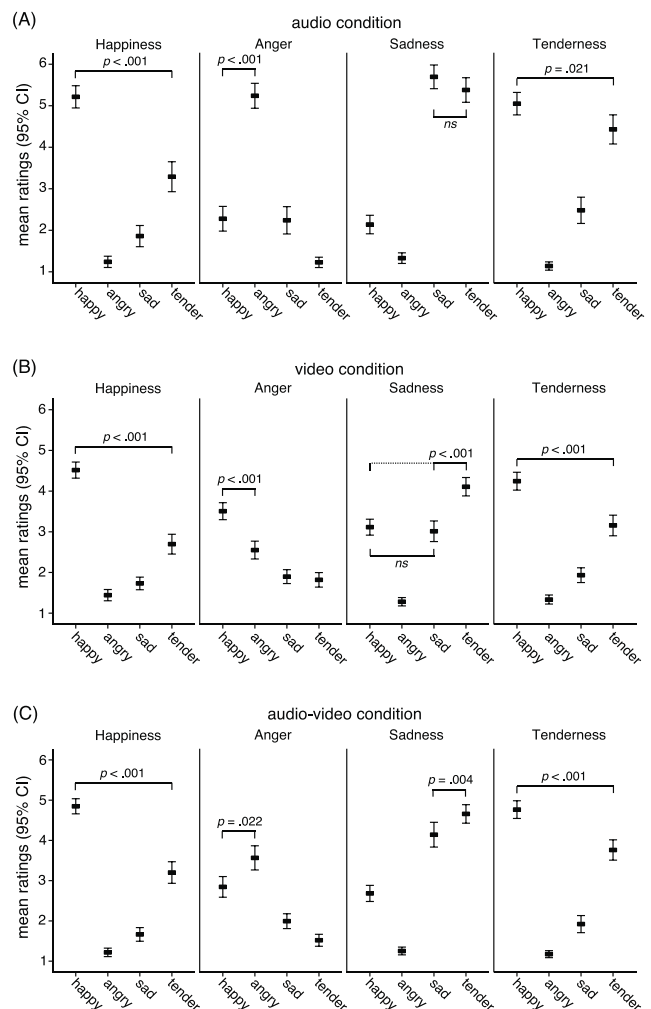
Subsequently, this section will present the analysis regarding which emotions were perceived in the different conditions, which condition was most effective in conveying the target emotion, and whether the perception of the audio-video clips was higher influenced by the auditory or by the visual modality of the clips.

### 3.1. Emotions perceived

In order to investigate which emotions were perceived in the different conditions and if the target emotions could be perceived, we displayed the ratings of the participants for the four musical stimuli (target emotion) as error bar plots (see Figure 1) and additionally conducted repeated measures ANOVAs (including subsequent post hoc tests) with the four ratings scales per stimulus to assess the significance of the differences between the ratings. The significance level of the post hoc tests (as indicated in Figure 1) was adjusted using Bonferroni correction to account for multiple comparisons.

Figure 1A shows the results for the four stimuli / target emotions of the audio condition. For the stimulus with Happiness as target emotion, the happiness scale received the highest average rating of the four scales/concepts, and the repeated measures ANOVA showed a significant main effect,  $F(2.34, 182.67) = 194.04$ ,  $p < .001$  (degrees of freedom adjusted using Greenhouse-Geisser correction as the sphericity assumption was violated – applied to all repeated measures ANOVA results, if not indicated differently).

The subsequent post hoc test revealed a highly significant difference ( $p < .001$ ) between happiness and tenderness, the second highest average, thus the target emotion Happiness



**Figure 1.** Error bar plots displaying average ratings and 95% confidence intervals for the four target emotion stimuli of the three conditions. (A) Audio condition. (B) Video condition, averaged across dancers. (C) Congruent audio-video condition, averaged across dancers.

could be successfully perceived in the audio condition. For the stimulus with the target emotion Anger, anger received the highest average rating of the four scales/concepts, the repeated measures ANOVA resulted in a significant main effect,  $F(2.07, 160.47) = 156.74$ ,  $p < .001$ , and the post hoc test showed a significant difference ( $p < .001$ ) between anger and happiness, the second highest average. Thus, the target emotion Anger could be successfully communicated as well. For the stimulus with Sadness as target emotion, sadness obtained

highest ratings with a significant main effect revealed by the repeated measures ANOVA,  $F(2.35, 183.27) = 310.84, p < .001$ . The post hoc test, however, indicated that the difference between sadness and the second highest average, tenderness, was non-significant ( $p = .68$ ), so there two emotions were slightly confused in case of the target emotion Sadness. For the stimulus with the target emotion Tenderness, happiness received the highest average rating of the four concepts, followed by tenderness. The repeated measures ANOVA exhibited a significant main effect,  $F(2.40, 187.29) = 171.86, p < .001$ , and the post hoc test resulted in a significant difference between both emotion concepts ( $p = .021$ ), so participants tended to confuse Tenderness with Happiness.

Figure 1B shows the results for the video condition. The ratings for the 16 clips presented in the experiment were averaged across the four dancers to obtain one rating per participant for each target emotion stimulus. For the target emotion Happiness, the happiness scale received the highest average rating of the four concepts, and the repeated measures ANOVA showed a significant main effect,  $F(2.36, 183.77) = 254.10, p < .001$ . The subsequent post hoc test revealed a highly significant difference between happiness and the second highest average, tenderness, ( $p < .001$ ), thus the target emotion Happiness could be communicated successfully in the video condition as well. For the target emotion Anger, happiness received the highest average rating of the four concepts, followed by anger. The repeated measures ANOVA indicated a significant main effect,  $F(2.21, 172.36) = 73.19, p < .001$ , and the post hoc test resulted in a significant difference between both concepts ( $p < .001$ ), so the movements performed to the stimulus rated as angry could not efficiently communicate anger. For the target emotion Sadness, tenderness received the highest average rating, followed by happiness and sadness. The repeated measures ANOVA revealed a significant main effect,  $F(2.33, 181.40) = 147.69, p < .001$ , and the post hoc comparison between sadness and tenderness as well as between tenderness and happiness exhibited significant differences ( $p < .001$ ), whereas the difference between sadness and happiness was shown to

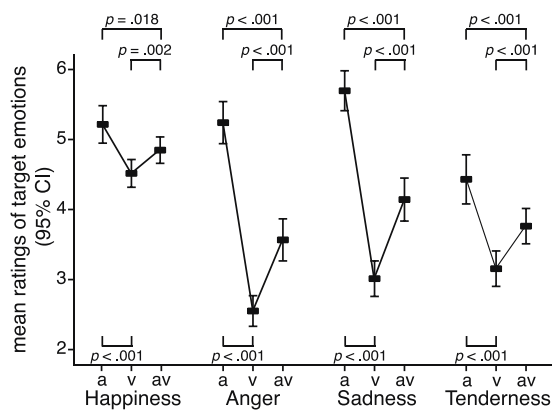
be non-significant ( $p = 1.00$ ). From this it can be concluded that the movements failed to communicate the intended emotion. For the target emotion Tenderness, happiness obtained the highest averaged rating, followed by tenderness. The repeated measures ANOVA resulted in a significant main effect,  $F(2.33, 181.63) = 198.98, p < .001$ , while the post hoc test indicated a significant difference between both concepts ( $p < .001$ ), so confusion alike to the audio condition occurred in the video condition as well.

Figure 1C displays the results for the congruent stimuli of audio-video condition. The ratings from the 16 congruent clips presented in the experiment were averaged across the four dancers to obtain one rating per participant for each target emotion stimulus. For the target emotion Happiness, happiness received the highest average rating of the four concepts, and the repeated measures ANOVA exhibited a significant main effect,  $F(2.38, 185.44) = 359.41, p < .001$ . The subsequent post hoc test revealed a significant difference between happiness and tenderness, the second highest average, ( $p < .001$ ), thus the target emotion Happiness could be communicated successfully in the audio-video condition. For the target emotion Anger, anger received the highest average rating of the four concepts. The repeated measures ANOVA resulted in a significant main effect,  $F(1.67, 130.41) = 60.76, p < .001$ , and the post hoc comparison between the two concepts showed a (moderately) significant difference ( $p = .022$ ), so the combination of auditory and visual components could fairly successfully communicate the target emotion Anger. For the target emotion Sadness, tenderness received the highest average rating, followed by sadness. The repeated measures ANOVA showed a significant main effect,  $F(1.88, 146.30) = 202.75, p < .001$ , and the post hoc test exhibited a significant difference between both concepts ( $p = .004$ ), which suggests that there occurred some confusion between the two. For the target emotion Tenderness, happiness obtained the highest average rating, followed by tenderness. The repeated measures ANOVA results revealed a significant main effect,  $F(2.18, 169.74) = 317.93, p < .001$ , and the post hoc comparison indicat-

ed a significant difference between the two concepts ( $p < .001$ ), so clips intended to convey tenderness communicated happiness instead. Thus, the same confusion as in the audio and video conditions described above happened in the audio-video condition.

### 3.2. Conveyance of target emotion

Next, we examined which experiment condition was most effective in conveying the four target emotions. Figure 2 illustrates the differences between the three conditions in conveying the four target emotions. For all target emotions, audio was most effective (highest average ratings), followed by the audio-video condition. The video condition was least effective in conveying the target emotions, especially in the case of anger and sadness. Repeated measures ANOVA results including post hoc tests (adjusted significance level using Bonferroni correction) revealed significant main effects (see Table 1) and differences between the three conditions in all target emotions (significance of the differences indicated in Figure 2).



**Figure 2.** Error bar plot showing the differences in mean ratings for the three experiment conditions with respect to conveying the four target emotions (a: audio condition; v: video condition; av: audio-video condition).

**Table 1.** Repeated measure ANOVA results for each target emotion.

	F statistics	significance
Happiness	$F(1.49, 116.13) = 14.03$	$p < .001$
Anger	$F(2, 156)^{\#} = 166.63$	$p < .001$
Sadness	$F(2, 156)^{\#} = 164.94$	$p < .001$
Tenderness	$F(1.54, 119.82) = 30.99$	$p < .001$

<sup>#</sup> no Greenhouse-Geisser correction applied (sphericity assumption holds)

### 3.3. Influence of modalities in audio-video condition

In order to investigate whether the perception of the audio-video clips was more strongly influenced by the auditory or by the visual modality, we conducted a series of linear regression analyses for each of the target emotions: 1) the audio ratings predicting the (congruent and incongruent) audio-video ratings, 2) the video ratings predicting the (congruent and incongruent) audio-video ratings, and 3) both the video and audio ratings predicting the (congruent and incongruent) audio-video ratings. The variances explained by each of the models ( $R^2$  values) are displayed in Table 2.

**Table 2.** Variances explained by each of the three regression models for each of the target emotions. The asterisks indicate the significance level of the regression models.

	$R^2$ (audio)	$R^2$ (video)	$R^2$ (audio + video)
Happiness	.54 ***	.39 ***	.93 ***
Anger	.82 ***	.14 **	.96 ***
Sadness	.62 ***	.33 ***	.94 ***
Tenderness	.51 ***	.39 ***	.91 ***

\*\*\*  $p < .001$ , \*\*  $p < .01$

For all four target emotions, the ratings for the audio condition could explain a larger amount of variance than the video ratings. We therefore assume that the participants paid more attention to the auditory part of the stimulus than to the visual part when rating the (partly incongruent) audio-video stimuli. Furthermore, audio and video ratings were additive, as adding both  $R^2$  values resulted in a value very close to the  $R^2$  value of the third (audio + video) regression model. Including both audio and video ratings in the (third) regression model, between 91 and 96 % of the total variance could be explained, which

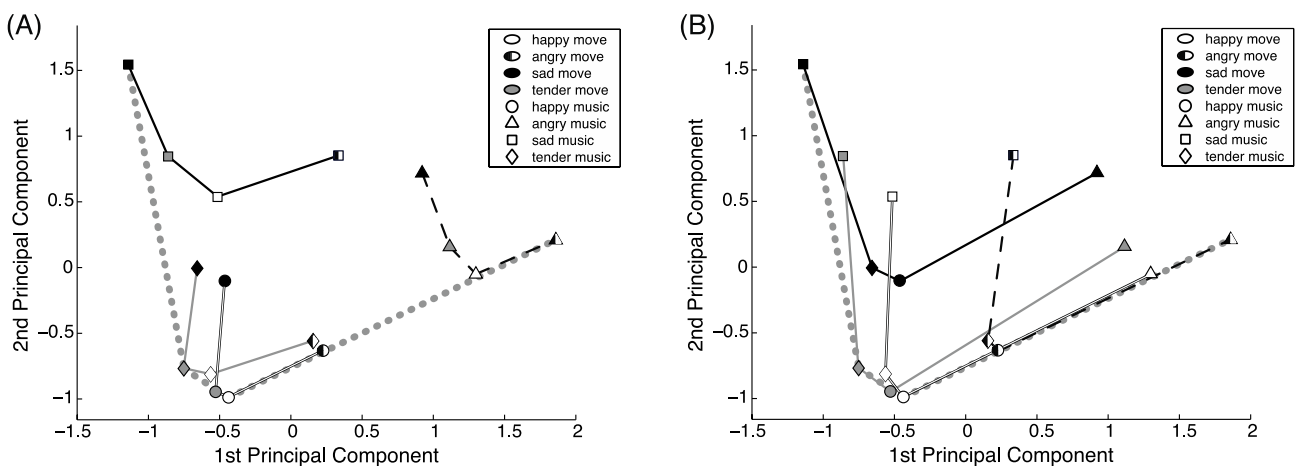
means that audio-video ratings could almost perfectly be predicted from both the ratings of the audio and the video conditions.

The final step of the analysis consisted of investigating the distribution and alignment of the ratings of the audio-video condition. To do so, we first reduced the dimensionality of the rating data for the audio-video condition using principal components analysis (PCA). Applying Kaiser's criterion, we retained two components that could account for 77.8% of the total variance (PC 1: 41.25% and PC 2: 36.55%). Subsequently, the components were rotated using varimax rotation and then averaged across dancers and participants yielding one value per audio/video combination resulting in 16 values per PC. Figure 3 displays these results as coordinates in a two-dimensional space. To show the influences of the auditory and visual modality on the audio-video ratings, the figure contains two subfigures: subfigure (A) shows the solution with the same audio stimuli being connected with lines, whereas subfigure (B) shows the same solution, but now with the same movement stimuli connected. The dotted line in each subfigure connects the congruent stimuli. Figure 3 shows that the stimuli having the same audio grouped closer together than the stimuli having the same movements. The happiness and tenderness cluster are overlapping in the same-audio-connected

emotion concepts. If we consider the two concepts as one cluster, we can see that there is no overlap between the happiness/tenderness cluster, the anger cluster, and the sadness cluster. In the same-movement case (Figure 3B) however, all clusters overlap, thus it seemed that the auditory domain was indeed dominating the visual domain in the audio-video condition. Furthermore, it is interesting to note that the clusters have self-similar structures with themselves and with the connectors of the congruent stimuli (dotted lines). This suggests that the ratings for the audio-video condition were very consistent and coherent, despite the contradictory information contained in the stimuli. This also accords with the additivity of the audio-only and video-only ratings in the regression models.

#### 4. Discussion

This study investigated whether quasi-spontaneous music-induced movements of non-professional dancers can convey emotional qualities. Participants were asked to rate audio, video, and audio-video clips according to the emotions expressed by music, movement, and combinations of both. The results showed that in general the target emotions could be perceived in the three experiment



**Figure 3.** Principal component solution averaged across dancers and participants plotted against each other in two dimensions. (A) Solution with the same audio stimuli being connected with lines. (B) Solution with the same movement stimuli being connected. The dotted line in each subfigure is connecting the congruent stimuli.

case (Figure 3A), which is in line with the previous results related to the confusion of the two

conditions, although systematic mismatches occurred, especially examples related to ten-



derness. Such confusions have been reported in the literature, both in music- and in movement-related research (e.g., Eerola & Vuoskoski, 2011; Gross et al., 2010; Pollick et al., 2001). Eerola and Vuoskoski (2011) reported that for the moderate examples used in their study on emotions perception in music happiness and tenderness as well as tenderness and sadness were confused, while Gross et al. (2010) found in investigating movements of emotional knocking that tenderness was confused with happiness. These findings suggest that it requires very stereotypical examples to communicate emotions, and more specifically to communicate a single emotion with one stimulus. The nature of the stimuli could serve as an explanation for the confusions related to tenderness. The prerequisite for our musical stimuli was to contain some kind of beat to induce and stimulate movement. However, it is difficult (if not impossible) to find musical stimuli that genuinely express tenderness and as well have a perceivable beat structure. Stimuli with a tender character, but possessing a beat are most likely to be rated as happy, as the activity/arousal level would increase due to the beat. Consequently, if the music communicated similar emotional qualities, it is likely that the movement exhibited to such music do so as well.

For the video condition, an interesting finding is that the movements performed to the angry stimulus could not communicate the same emotion. An explanation for this finding could be that dance movements are commonly understood as positive and pleasant (leisure activity, fun), so they might appear and thus be rated (more) positively as well. This explanation would connect the discrete emotions approach with the dimensional model of emotions (Russell, 1980), in which happiness is commonly described as active positive/pleasant, anger as active negative, sadness as inactive negative, and tenderness as inactive positive, so the shift from anger to happiness could be explained in terms of the valence/pleasantness dimension of that model. Additionally, the movements might contain less stereotypical hints to emotional qualities than the music, at least to negative emotions. Thus, it could be that negative emotions can

only be communicated in dance movements, if the dancers are asked to move according to the emotional character of the music (and not "just" to the music, as it was in this case). Similar results were also found in studies on dance and on acted emotions (Atkinson et al., 2004; Dittrich et al., 1996), although they would disagree with Walk and Homan's (1984) 'alarm hypothesis': expressions of negative emotions are easier and better recognized than positive emotions. The tendency towards more positive judgments (related to the dimensional model of emotions) can also be found in the case of Sadness as target emotion. The sad examples were rated being high in expressing tenderness, so the observers rated them more positive and pleasant, like they did with the examples expressing anger. A similar shift might have occurred in the ratings for the target emotion Tenderness when being confused with happiness, though it would be rather related to the activity level than to pleasantness in terms of the dimensional model of emotions.

We observed similar shifts in the congruent audio-video condition for the Sadness and Tenderness as target emotions – sad examples being perceived as more positive (cf., dimensional model) and tender examples being perceived as more active – though the results for the target emotion anger were different: anger could be successfully communicated when audio and video was presented together. Thus, it seems that the auditory domain dominates such ratings compared to the visual domain.

A closer investigation of the question which condition was most effective in conveying emotions revealed that the audio condition received the highest average ratings for the four target emotions, followed by the audio-video condition. For all target emotions, the same pattern emerged. Thus, we failed to find support to our initial hypothesis that the audio-video condition would receive highest recognition. We could neither confirm the results obtained in previous research (Collignon et al., 2008; Davidson, 1993; Petrini et al., 2010; Sörgjerd, 2000; Vines et al., 2006). It could be that music is the most powerful carrier of emotions, at least in relation to spontaneous dance movements. As mentioned already, dance movement might rather express

pleasure, sensuality, and aesthetics, so they might lack stereotypical and stylized emotion-related qualities.

Earlier studies usually employed professional actors or dancers, whose task was to portray specific emotions. Such an approach leads to rather stylized movements that might successfully express emotions, but may lack naturalism. Our stimuli, on the other hand, were derived from quasi-spontaneous movements to music, so they might lack emotion-specific qualities, as the dancers were not instructed to express emotions with their movements. However, although the dancers were unaware of any implications to emotions, they still moved in a way that observers perceived as expressive of emotions.

The three experiment conditions showed a noteworthy relationship, as the auditory and visual modalities appeared to be additive with regards to the audio-video presentation of the stimuli. The audio-video ratings could be almost perfectly predicted from both the ratings for the audio condition and the video condition, whereas the ratings of either the audio condition or the video condition could only partly predict the audio-video ratings. This would suggest that both domains influence the perception of audio-video stimuli, and that participants tried to integrate both channels, as has already been shown in research on music (Petrini et al., 2010; Vines et al., 2006) and on voice-face integration (De Gelder & Vroomen, 2000; Massaro & Egan, 1996).

Furthermore, we found that in the audio-video condition, the auditory modality dominated the visual modality. Thus, regardless of the combination of the stimuli (i.e., for both congruent and incongruent combinations of music and movements), participants seemed to rather rate according to the music than to the movement. This result supports our initial hypothesis as well as previous studies in music research (Petrini et al., 2010; Vines et al., 2006), but contradicts with results from research on voice-face stimuli (Collignon et al., 2008) – a finding that would suggest that differences exist in musical and linguistic processing with regards to emotion perception. The domination of the auditory domain could mean that the music was more expressive of a certain

emotion than was the movement, so it was guiding the ratings of the participants – especially in cases of incongruent combinations. This issue will be investigated further in the future, as data was gathered about the participants' own perception of their rating behavior: the participants were asked after completion of the audio-video part to indicate whether they paid more attention to the music, to the movement, or if they tried to integrate both. These results might give us more insight into the participants' integration of auditory and visual stimulus information.

An interesting finding within the results of the audio-video condition is that the alignment of the ratings showed a self-similar structure. We plotted the principal component solution of the basic emotion ratings in a two-dimensional space and connected the stimuli in two ways: related to the examples containing the same music and related to the examples containing the same movements. The resulting clusters supported the already mentioned finding about the auditory domain having higher influence on the ratings, as the same-music clusters are non-overlapping and smaller than the same-movement clusters. It seems that the music or/and the congruent stimulus combination "attracted" the other movements to cluster close together. Furthermore, the clusters exhibited self-similar structures with each other and with the cluster of the congruent stimuli. This suggests that, despite the contradictory information of the stimuli, the ratings were very consistent and coherent. It could be assumed that both the congruent stimulus combinations and the music were somehow guiding the perception of the incongruent stimulus combinations, so the participants followed a homogeneous rating scheme throughout the experiment.

The result that the movements failed to communicate anger in the video condition would deserve further investigation. In a future experiment, participants could be asked – differently from the present study – to express, while dancing, the emotion expressed in the music. It would be of interest to investigate how the movements (and their perception) change and if participants are able to maintain

dance movement characteristics while successfully communicating emotions.

The selection of the musical stimuli in this experiment was somewhat restricted, insofar as they were derived/adopted from a previous study. Therefore, the stimuli might not have as clearly expressed an emotion as stimuli selected purely based on their emotional characteristics. However, a follow-up study could address this issue with a fresh set of stimuli of more clearly expressing an emotion, though being potentially less danceable than the recent selection.

The presented study gave revealing insights into multimodal perception of dance movement stimuli showing that participants could attribute emotions to dance movements. They were furthermore able to meaningfully integrate audio and video signals, even in case of incongruent combinations. This investigation was a follow-up study to a motion capture study (Burger, Saarikallio, et al., 2013), in which we could establish links between movement features and the emotional characteristics of music. The results of both studies can serve as support for each other: we could show in both studies that music-induced movements can express emotional characteristics, with the computational analysis showing in particular that there are relationships between movement features and the emotional content of music, and the perceptual experiment showing that humans can perceive and recognize emotions in music-induced movements.

## Acknowledgments

This study was funded by the Academy of Finland (projects 118616, 125710, 136358).

## 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.

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, 17(1), 43–64.

Boone, R. T., & Cunningham, J. G. (1998). Children's Decoding of Emotion in Expressive Body Movement: The Development of Cue Attunement. *Developmental Psychology*, 34(5), 1007–1016.

Burger, B., Saarikallio, S., Luck, G., Thompson, M. R., & Toiviainen, P. (2013). Relationships between perceived emotions in music and music-induced movement. *Music Perception*, in press.

Burger, B., Thompson, M. R., Saarikallio, S., Luck, G., & Toiviainen, P. (2013). Influences of rhythm- and timbre-related musical features on characteristics of music-induced movement. *Frontiers in Psychology*, 4:183.

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain research*, 1242, 126–35.

Dahl, S., & Friberg, A. (2007). Visual Perception of Expressiveness in Musicians' 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.

De Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion*, 14(3), 289–311.

De Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, 13(4), 247–268.

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.

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39(1), 18–49.

Evans, P., & Schubert, E. (2008). Relationships between expressed and felt emotions in music. *Musicae Scientiae*, 12(1), 75–99.

Gabrielsson, A. (2002). Emotion perceived and emotion felt: same or different? *Musicae Scientiae*, 6, 123–147.

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., & Lindström, E. (2010). The role of structure in the musical expression of emotions. In P. N. Juslin & J. Sloboda (Eds.), *Handbook of*

*Music and Emotion. Theory, Research, Applications* (pp. 367–400). Oxford, UK: Oxford University Press.

Gross, M. M., Crane, E. a., & Fredrickson, B. L. (2010). Methodology for Assessing Bodily Expression of Emotion. *Journal of Nonverbal Behavior*, 34(4), 223–248.

Iverson, J. M., & Thelen, E. (1999). Hand, Mouth and Brain. *Journal of Consciousness Studies*, 6(11-12), 19–40.

Krumhansl, C. L. (2002). Music: A Link Between Cognition and Emotion. *Current Directions in Psychological Science*, 11(2), 45–50.

Lagerlöf, I., & Djerf, M. (2009). Children's understanding of emotion in dance. *European Journal of Developmental Psychology*, 6(4), 409–431.

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. Cambridge, MA, London, UK: MIT Press.

Leman, M., & Godøy, R. I. (2010). Why Study Musical Gesture? In R. I. Godøy & M. Leman (Eds.), *Musical Gestures. Sound, Movement, and Meaning* (pp. 3–11). New York, NY: Routledge.

Luck, G., Saarikallio, S., Burger, B., Thompson, M. R., & Toiviainen, P. (2010). Effects of the Big Five and musical genre on music-induced movement. *Journal of Research in Personality*, 44(6), 714–720.

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review*, 3(2), 215–221.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press.

Montepare, J. M., Goldstein, S. B., & Clausen, A. (1987). The identification of emotions from gait information. *Journal of Nonverbal Behavior*, 11(1), 33–42.

Petrini, K., McAleer, P., & Pollick, F. (2010). Audiovisual integration of emotional signals from music improvisation does not depend on temporal correspondence. *Brain research*, 1323, 139–48.

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

Russell, J. A. (1980). A Circumplex Model Of Affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178.

Saarikallio, S., Luck, G., Burger, B., Thompson, M. R., & Toiviainen, P. (2013). Dance moves reflect current affective state illustrative of approach-avoidance motivation. *Psychology of Aesthetics, Creativity, and the Arts, in press*.

Schubert, E. (1999). Measuring Emotion Continuously: Validity and Reliability of the Two-Dimensional Emotion-Space. *Australian Journal of Psychology*, 51(3), 154–165.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428.

Sörgjerd, M. (2000). *Auditory and visual recognition of emotional expression in performances of music*. Unpublished thesis. Department of Psychology, University of Uppsala, Sweden.

Toiviainen, P., & Burger, B. (2013). *MoCap Toolbox Manual*. University of Jyväskylä: Jyväskylä, Finland. Available at <http://www.jyu.fi/music/coe/materials/mocaptoolbox/MCTmanual>.

Van Dyck, E., Moelants, D., Demey, M., Deweppe, A., Coussement, P., & Leman, M. (2013). The Impact of the Bass Drum on Human Dance Movement. *Music Perception*, 30(4), 349–359.

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.

Walk, R. D., & Homan, C. P. (1984). Emotion and dance in dynamic light displays. *Bulletin of the Psychonomic Society*, 22, 437–440.

Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, 28(6), 879–896.

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion*, 8(4), 494–521.