Cognition 118 (2011) 157–170

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance

Bradley W. Vines [a,c,d,*], Carol L. Krumhansl [b], Marcelo M. Wanderley [c], Ioana M. Dalca [a], Daniel J. Levitin [a,c]

[a] Psychology Department, McGill University, Canada
[b] Psychology Department, Cornell University, USA
[c] Schulich School of Music and Center for Interdisciplinary Research in Music Media and Technology, McGill University, Canada
[d] Institute of Mental Health, Department of Psychiatry, University of British Columbia, Canada

ARTICLE INFO

Article history:
Received 30 January 2010
Revised 15 November 2010
Accepted 16 November 2010
Available online 10 December 2010

Keywords:
Cross-modal interaction
Multi-sensory integration
Music cognition
Emotion
Performance
Expressivity
Gesture
Non-verbal communication

ABSTRACT

We investigate non-verbal communication through expressive body movement and musical sound, to reveal higher cognitive processes involved in the integration of emotion from multiple sensory modalities. Participants heard, saw, or both heard and saw recordings of a Stravinsky solo clarinet piece, performed with three distinct expressive styles: restrained, standard, and exaggerated intention. Participants used a 5-point Likert scale to rate each performance on 19 different emotional qualities. The data analysis revealed that variations in expressive intention had their greatest impact when the performances could be seen; the ratings from participants who could only hear the performances were the same across the three expressive styles. Evidence was also found for an interaction effect leading to an emergent property, intensity of positive emotion, when participants both heard and saw the musical performances. An exploratory factor analysis revealed orthogonal dimensions for positive and negative emotions, which may account for the subjective experience that many listeners report of having multi-valent or complex reactions to music, such as "bittersweet."

1. Introduction

By investigating the perception of musical performance, we can explore fundamental processes in human communication spanning non-verbal music and speech. Extensive research has examined relations between speech and those paralinguistic gestures that accompany speech (e.g., facial expressions, hand gestures, and postures), as well as the effect of these modalities of human expression on the observer's experience (Goldin-Meadow, 2003; McNeill, 2005). Like language, music is a complex, natural human behavior that can be found in all human cultures. With regard to human evolution, music might predate full spoken language (Fitch, 2006; Mithen, 2007). Therefore, studying the perception of musicians' physical gestures, as they relate to the musical sound, complements research into speech–gesture relations, and has the potential to reveal general processes involved in inter-personal communication.

Assuming that music constitutes a form of communication (Bharucha, Curtis, & Paroo, 2006; Jackendoff & Lerdahl, 2006; Meyer, 1956), audience members and their manner of observing music must play an essential role. Lasswell's (1964) communication theory can be characterized by the question, "Who says what, in what channel, to whom, and with what effect," and is clearly relevant to musical performance, at least for western concert music. Cross-cultural factors, such as the listener's familiarity with the

* Corresponding author. Address: Institute of Mental Health, Department of Psychiatry, University of British Columbia, 430-5950 University Blvd., Vancouver, BC, Canada V6T 1Z3. Tel.: +1 604 827 4364; fax: +1 604 827 3373.
E-mail address: bradley.vines@ubc.ca (B.W. Vines).

0010-0277/$ - see front matter
doi:10.1016/j.cognition.2010.11.010

music's tonal system, play an important role as well (Balkwill & Thompson, 1999). The question of what perceptual channel emotions in music pass through, from performer to listener, remains unresolved. Musical emotion may be derived solely from the sounds of a musical work, and may be mediated by schematic expectations and knowledge of musical forms (Davies, 1994; Lerdahl, 2001; Levinson, 1980). On the other hand, the physical gestures of musicians may also contribute to the transduction of musical emotion, by means of the visual modality in addition to the sound (Lopes & Bergeron, 2009; Vines, Krumhansl, Wanderley, & Levitin, 2006); indeed, the pianist Glenn Gould's body movements changed as a function of whether or not an audience was present (Delalande, 1988). Even when the performer cannot be seen, the listener's brain may process music in terms of the body movements from which the sounds originate (Galati et al., 2008; Gazzola, Aziz-Zadeh, & Keysers, 2006; Hauk, Shtyrov, & Pulvermuller, 2006).

The influence of visual information on auditory perception is well established (McGurk & MacDonald, 1976), and even tactile input has been shown to influence speech perception (Gick & Derrick, 2009). Such cross-modal demonstrations are typically restricted to the perception of low-level information, such as phonemes. Recently, however, studies have begun to show the influence of visual information on *emotional* perception and cognition of auditory events in music (Chapados & Levitin, 2008; Davidson, 1993, 1994; Vines et al., 2006). Gabrielsson (2001) further suggests that the experience of emotion in music involves an interaction between musical, personal, and situational factors. That is, musical emotion results from the relation between the *sounds* of the music and factors that are observer-specific, such as the observer's state of attentiveness, and whether or not the performer can be seen.

The specific content that music conveys also remains a conundrum. Music, like language, embodies a coherent structure, the meaning of which unfolds over time with constituent elements arranged hierarchically (Cooper & Meyer, 1960; Krumhansl, 1990; Lerdahl & Jackendoff, 1983; Levitin & Menon, 2003, 2005; Patel, 2003). However, whereas spoken language generally *refers* to specific objects, events, or ideas, music represents the abstract dynamics of emotional life without indexing specific referents in the world (Brown, 2000; Cross, 2003). Evidence does suggest that instrumental music can activate semantic–linguistic networks in the brain. For example, the sound of very high violin notes primes the word *needle* (Koelsch et al., 2004). Many people around the world invest significant resources in listening to, learning about and performing music – in North America, people spend more money on music than on prescription drugs or sex (Huron, 2001). Perhaps this is because music can activate deep neural structures associated with reward, pleasure and the mediation of dopaminergic levels (Blood & Zatorre, 2001; Menon & Levitin, 2005), can elicit emotions that are as vivid and intense as "real-world" emotions associated with life events (Gabrielsson & Lindström, 2003), and can initiate such experiences in less than 1 s (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005). The vividness and intensity of emotions elicited by music may result from an interaction between the perception of musical events, and expectations about the music based upon general principles of perceptual organization and cultural influences (Krumhansl, 2002). But what is the structure of musically-induced emotions, and is it similar to or different from the structure of real-world emotions?

1.1. Relevant emotion research

An issue in the study of emotion that is particularly relevant to the structure of musically-induced emotions involves the relation between positive and negative valence. Those who support the dimension approach for characterizing emotions assume that emotions do not have distinct prototypes, but differ from one another along a set of dimensions. A central question within the dimension literature concerns how positive and negative valence are related. Arousal–valence theory and evaluative-space theory encapsulate the two principal perspectives on this issue.

1.2. Arousal–valence theory

The arousal–valence theory or *circumplex model*, developed by (Russell, 1979, 1980, 2003; Russell & Carroll, 1999), treats all emotions as if they fall into a two-dimensional representation involving valence (positive to negative) and arousal (very awake to asleep). Within the arousal–valence framework, appetitive (positive valence) and aversive (negative valence) experiences occur at opposite ends of the same bi-polar dimension and are, therefore, mutually exclusive (e.g., the experience of positive affect negates the experience of negative affect). Several studies support the arousal–valence model (Faith & Thayer, 2001; Green, Goldman, & Salovey, 1993; Green & Salovey, 1999), which can account for emotional responses to a variety of stimuli. For example, Schubert (1999, 2004) employed the arousal–valence model to study real-time emotional responses to musical stimuli.

1.3. Evaluative-space theory

Another model, based upon the evaluative-space theory (Cacioppo & Berntson, 1994; Cacioppo, Gardner, & Berntson, 1997), can represent multi-valent emotional states. Positive valence and negative valence occupy separate, orthogonal dimensions in the model. Larsen, McGraw, and Cacioppo (2001) found evidence supporting this model under atypical circumstances involving complex social influences. For example, while student participants were moving out from their dorms for summer vacation, graduating from a university, or viewing an emotional film, there was a much higher incidence of multi-valent emotional states in which both positive and negative experiences coexisted. Additionally, Larsen, Norris, and Cacioppo (2004) identified emotional reactions to issues of social importance that involved multi-valent emotions, with both positive and negative aspects occurring at once (e.g., a person's emotional response to the question of whether there should be capital punishment). Other research groups have also found evidence for orthogonal dimensions of positive and negative valence (Watson &

Tellegen, 1985), and Hunter, Schellenberg, and Schimmack (2008) found that music in particular could evoke mixed happy and sad emotional responses, as well as mixed pleasant and unpleasant responses.

These models may prove useful in determining the structure of musically-induced emotions, and how musicians communicate emotion through their performances.

1.4. Expression of emotion in music

Musicians use a mixture of auditory cues, body movements, and practice methods in their efforts to communicate emotion through musical performance (Gabrielsson, 1999). Important auditory cues include tempo changes (e.g., accelerando, decelerando), loudness dynamics, vibrato, and note asynchrony (which is especially relevant for piano performance; Repp, 1996). Although musicians' body movements are generally unintended (Wanderley, 2002) – as are paralinguistic movements that accompany speech (McNeill, 1992, 1999) – such movements in general do convey information about performers' mental states, including their expressive intentions and emotions (Davidson, 1993; Dittrich, Troscianko, Lea, & Morgan, 1996; Hatfield, Cacioppo, & Rapson, 1994; Runeson & Frykholm, 1983). Musicians' body movements, such as head, eyebrow, and postural adjustments, significantly reinforce, anticipate, or augment the content in sound at structurally important points in the performance (Vines, Nuzzo, & Levitin, 2005; Vines et al., 2006), and can influence perceived dissonance and valence in the music (Thompson, Graham, & Russo, 2005). Musicians can systematically manipulate these various cues to alter the expressive qualities of their music, including the interpretation of phrasing structure, stylistic modifications, and emotional content. For example, singers' facial expressions can influence the emotion conveyed through sound (Thompson, Russo, & Quinto, 2008). Even in the absence of sound, it is possible to identify a performer's emotional intentions simply by watching videos of the performance (Dahl & Friberg, 2007). Musicians' movements can also affect the perception of low-level musical properties, such as timbre, loudness, pitch, and note duration (Schutz, 2008). For example, the movements of a marimba player can influence the perception of auditory duration for listeners who view the performance (Broughton, Stevens, & Malloch, 2009; Schutz & Lipscomb, 2007).

A common pedagogical tool used in musical training to hone control over emotive expression involves the use of different performance manners to emphasize particular aspects of sound production and communication (Davidson & Correia, 2002). The restrained manner focuses a musician's attention upon technique. The standard manner emphasizes naturalness. And the exaggerated manner calls for enhanced musical dynamics and emotion. For the scientist, these three performance manners are useful for understanding how a performer's expressive intentions influence sound and body movement, and through what sensory modality performers convey their intentions to audience members.

Davidson (1993, 1994) found that seeing a performance could open unique pathways of communication between performer and audience. For these studies, Davidson recorded musicians with audio and synchronized video as they played musical segments using three different levels of expressiveness. Later, experimental participants saw, heard or both saw and heard the performances with the visual aspect presented in point-light form (after Johansson, 1973). The participants judged the expressivity of each performance on a Likert scale. Results showed that the participants were best able to distinguish between the three levels of intended expressiveness when the performer could be seen, *even if not heard*, a trend which held true for both musician and non-musician observers (Davidson, 1993, 1994). These findings provide preliminary evidence that the visual aspect of musical performances augments information in sound by revealing performers' musical intentions. However, the question of what emotional qualities are conveyed through seeing and hearing musical performances remains unanswered.

In the present paper, we address the following questions: through which sensory channels do musical performers' expressive intentions affect the observer's emotional response, and what is the structure of the experience of emotion that they communicate? We explored these questions by investigating the emotional impact of musical stimuli that varied in terms of the level of expressive intention, and whether participants saw, heard, or both saw and heard the performances. We sought to preserve ecological validity by using a piece of music in the standard repertoire by a major composer, and recordings of live performances as stimuli.

2. Methods

2.1. Participants

Thirty participants from the McGill University community were recruited (mean age = 23.7 years, range = 18–30, SD = 3.1). All reported at least 5 years of musical training (M = 13.5 years, range = 5–26, SD = 6.2). Previous research found that musicians and non-musicians tended to make similar judgments of emotion and segmentation in response to a piece of music, and that there tended to be less variance in the judgments of musicians compared to the judgments of non-musicians (Deliège & El Ahmade, 1990; Fredrickson, 2000; Krumhansl, 1996). In pilot testing, we found that judgments made by non-musicians were similar to those made by musicians for the task used in this experiment. This suggests that results based upon the judgments of musician participants in this study may be considered representative of non-musicians as well. Participants received five Canadian dollars for taking part in the study.

Participants were randomly assigned to one of three treatment groups of 10 participants each. The auditory only (AO) group only heard the performances, the visual only (VO) group only saw the performances, and the auditory + visual (AV) group both heard and saw the performances. We used a between-subjects design for presentation condition in order to avoid the possibility of a subject in the AO condition being able to imagine the movements of the musicians while listening, or a subject in the VO condition being able to imagine the sound of the piece while seeing the performance. This approach decreased the power of our analysis, but eliminated any potential carry-over effects due to experiencing the same performances in all three presentation conditions.

2.2. Stimuli

The stimuli were audio–video recorded performances by two professional clarinetists. The clarinetists played Stravinsky's *Second Piece for Clarinet Solo* (Revised edition 1993, Chester Music Limited, London) with three different performance manners: restrained (playing with as little body movement as possible), standard (naturally, as if presenting to a public audience), and exaggerated (with enhanced expressivity). There were six videos in total (two performers, and three performance manners) ranging in duration from 68 to 80 s with a mean duration of 73 s, and a standard deviation of 5 s. These video recordings were used in previous studies by the authors: Wanderley (2002) investigated the performers' movement patterns using Optotrak recordings; Vines and colleagues (2006) investigated the contribution of auditory and visual information to the real-time perception of phrasing and tension in the performances.

We prepared the stimuli such that approximately 1 s of video preceded the onset of the first note, or one second of silence in the case of the AO stimuli. We created a computer program in the Max programming environment (Cycling '74, San Francisco, CA, USA), on a Macintosh G4 (Apple Computer Company, Cupertino, CA, USA) with a Mitsubishi 20 in. Flat Panel LCD monitor to present instructions to the participants, to present the stimuli, to collect data, and to interface with the participants. The computer program presented the performances in a new random order for each participant, with digital-video quality (16-bit Big Endian codec, 720×576 pixels), and National Television Standards Committee format (NTSC 25 frames per second). The participants in the AO and AV conditions listened to the performances over Sony MDR-P1 headphones, and adjusted the volume to a comfortable level. For consistency across conditions, we asked participants in the AO condition to keep their eyes open, and to look at the black computer screen while they listened to the performance. Participants in the VO condition also wore the same headphones during the stimulus presentation, though there was no sound.

2.3. Task

Following each of the six performances, participants made Likert-scale ratings for each of 19 words. (There was no training on the task or prior exposure to the stimuli.) Participants made their responses with the computer mouse by selecting one of five buttons appearing in a horizontal row on the monitor. The buttons were numbered from 1 (*not at all*) to 5 (*very much*), with a left-to-right orientation. Similar 5-point scales have been used in previous research on emotion (Faith & Thayer, 2001). The words appeared in a new random order for each performance and for each participant, along with the following instruction:

> *Rate how much you yourself experienced the following sensations during that last performance.*

We collected judgments for the following words: amusement, anger, anxiety, contempt, contentedness, disgust, embarrassment, expressivity, familiarity, fear, happiness, intensity, interest, movement, pleasantness, quality, relief, sadness, and surprise. We drew these words from the literature on emotion research and music cognition (Eibl-Eibesfeldt, 1972; Ekman, 1992, 1998; Izard, 1971; Krumhansl, 1997; Krumhansl & Schenck, 1997; McAdams, Vines, Vieillard, Smith, & Reynolds, 2004; Ortony & Turner, 1990; Russell, 1979).

2.4. Analysis

The data were analyzed with a repeated-measures ANOVA, a between-subjects factor "mode of presentation" (AO, VO, AV), and within-subjects factors "performer" (performer W, performer R), "manner" (restrained, standard, exaggerated), and "emotion" (the 19 words mentioned above).

We also entered all the data (3420 judgments from 19 words $\times$ 30 participants $\times$ 6 stimuli) into a factor analysis to identify the primary orthogonal modes of variation in the data. We then applied a repeated-measures ANOVA to the factor scores from each primary mode of variation. These ANOVAs had a between-subjects factor "mode of presentation," and within-subjects factors "performer" and "manner."

3. Results and discussion

3.1. ANOVA of all judgments

The repeated-measures ANOVA on all judgments yielded main effects for the factors mode of presentation, manner, and emotion.

Regarding the main effect for the between-subjects factor mode of presentation ($F(2, 27) = 4.77$, $p = .02$), a post hoc comparison with a Holm–Bonferroni correction at the .05 significance level revealed that, overall, the emotion ratings were significantly lower for the VO group compared to both the AO group and the AV group; the AO and AV groups did not differ ($p = .90$, before correction for multiple comparisons). This result provided evidence that the emotional intensity conveyed through sound was significantly greater than that conveyed through the visual modality in the musical performances. However, there may have been situational variables that contributed to this result. For example, familiarity with the presentation condition was certainly different for the VO group compared to the AO and AV groups. It is common to experience (and to engage emotionally with) music in sound alone, or in a combination of sound and video. It is not surprising, therefore, that participants in the VO condition made emotion ratings of a lower intensity, being that they were relatively unfamiliar with experiencing music in that way, and aware that they were missing the audio component, which under normal circumstances is the primary modality in music.

The main effect for the factor manner ($F(2, 54) = 12.70$, $p < .001$) was driven by lower ratings overall in response to the restrained performances. The restrained performances elicited significantly lower emotion ratings compared to both the standard performances and the exaggerated performances, according to a post hoc comparison with a Holm–Bonferroni correction at the .05 significance level; the responses to the standard and exaggerated performances did not differ ($p = .34$, before correcting for multiple comparisons). Overall, the restraint of movement dampened the emotional intensity conveyed to observers, particularly for the VO and AV modes of presentations.

There was also a significant main effect for the emotion factor ($F(18, 486) = 48.16$, $p < .001$). Fig. 1 shows the mean and standard deviations for the overall ratings of each word. The musical composition itself likely determined which emotions would be rated most highly. A different piece that was slow moving and melancholy in sound, for example, might have led to high ratings for words like *sadness*. The Stravinsky piece, however, was best characterized by words like *expressivity* and *movement*. This may be due to the relatively high note density in the piece, the wide frequency range covered by the notes, and the wide dynamic range.

There was a significant interaction effect between the factors manner and emotion ($F(36, 972) = 3.91$, $p < .001$), and a significant three-way interaction for the factors mode of presentation, manner, and emotion ($F(72, 972) = 1.82$, $p < .001$). We did the following analyses to understand these interaction effects.

The first analysis compared the emotion ratings for each performance manner, within the three modes of presentation. We used 2-tailed paired-samples *t*-tests and a Holm–Bonferroni correction for multiple comparisons with a significance level of .05. Data for both performers were combined in each *t*-test. Fig. 2 shows the results of the analysis. These tests revealed that differences between performance manners only affected judgments in the VO and AV modes of presentation.

We performed a second analysis to reveal the effect of being able to see the performances. This analysis compared the emotion ratings for the AO and AV modes of presentation, within the three performance manners. We used 2-tailed independent-samples *t*-tests, with a Holm–Bonferroni correction for multiple comparisons with a significance level of .05. Data for both performers were combined in each *t*-test. Fig. 3 shows the results of this analysis. A notable finding was that the AV ratings of *happiness* for the exaggerated performance manner were significantly higher than the AO ratings. This provided evidence for an emergent intensity of positive emotion when participants could see the performances in addition to hearing them.

The interaction effect between mode of presentation and emotion did not reach significance ($p = .13$).

3.2. Factor analysis

The meanings of the words in this study overlapped to some extent. For example, the emotions *anger* and *contempt* shared common characteristics. Judgments for emotions with similar meanings were likely to be correlated. We applied a factor analysis to identify the most prominent orthogonal dimensions of emotion. The exploratory factor analysis (as implemented in SPSS 11) comprised a principal components analysis and varimax rotation.

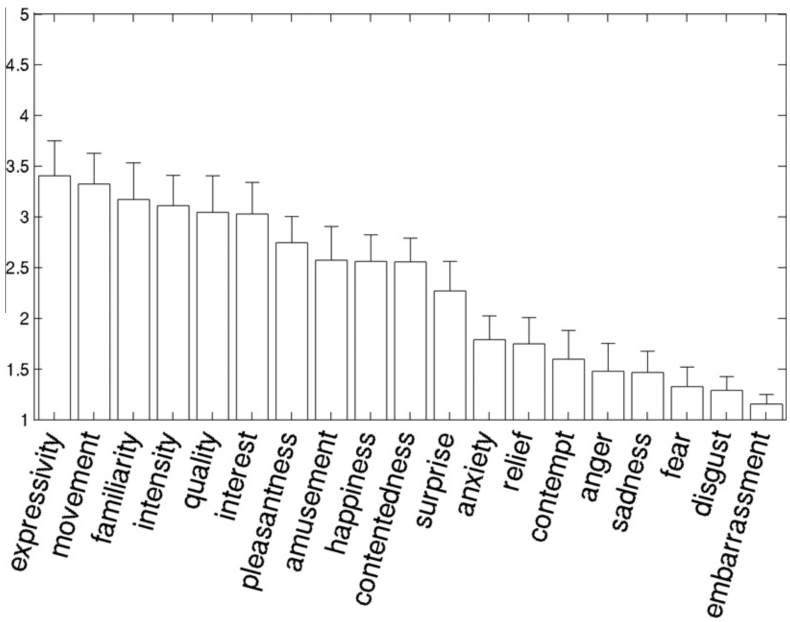

Fig. 1. Mean values and error bars showing the 95% confidence interval for ratings of each word. The performances were most characterized by active and positive words, and least by negative words.

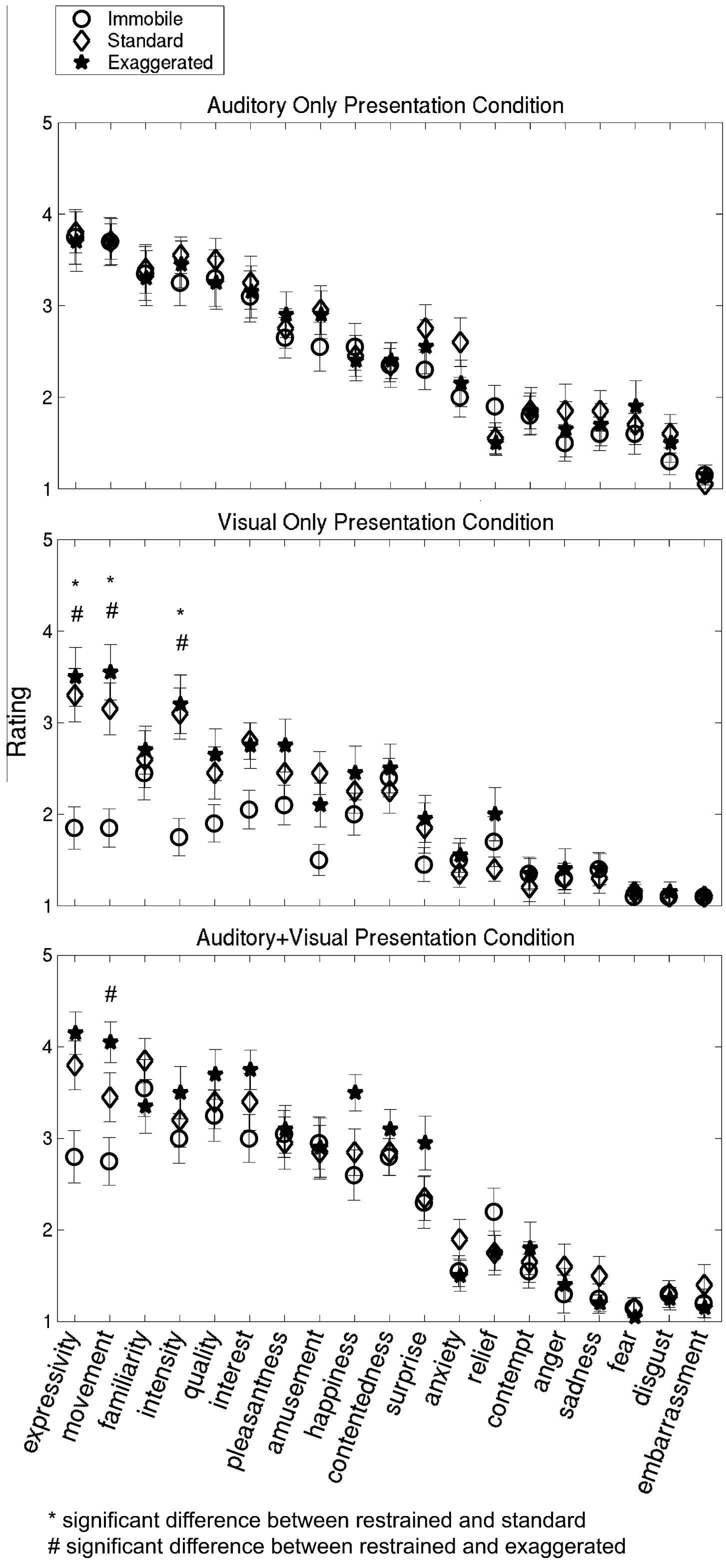

Fig. 2. Results are shown separately for the three modes of presentation. Symbols indicate significant differences at the $p < .05$ level, as noted at the bottom of the figure. The variations in performance manner did not effect judgments made by subjects in the AO mode of presentation. Error bars show the standard error mean.

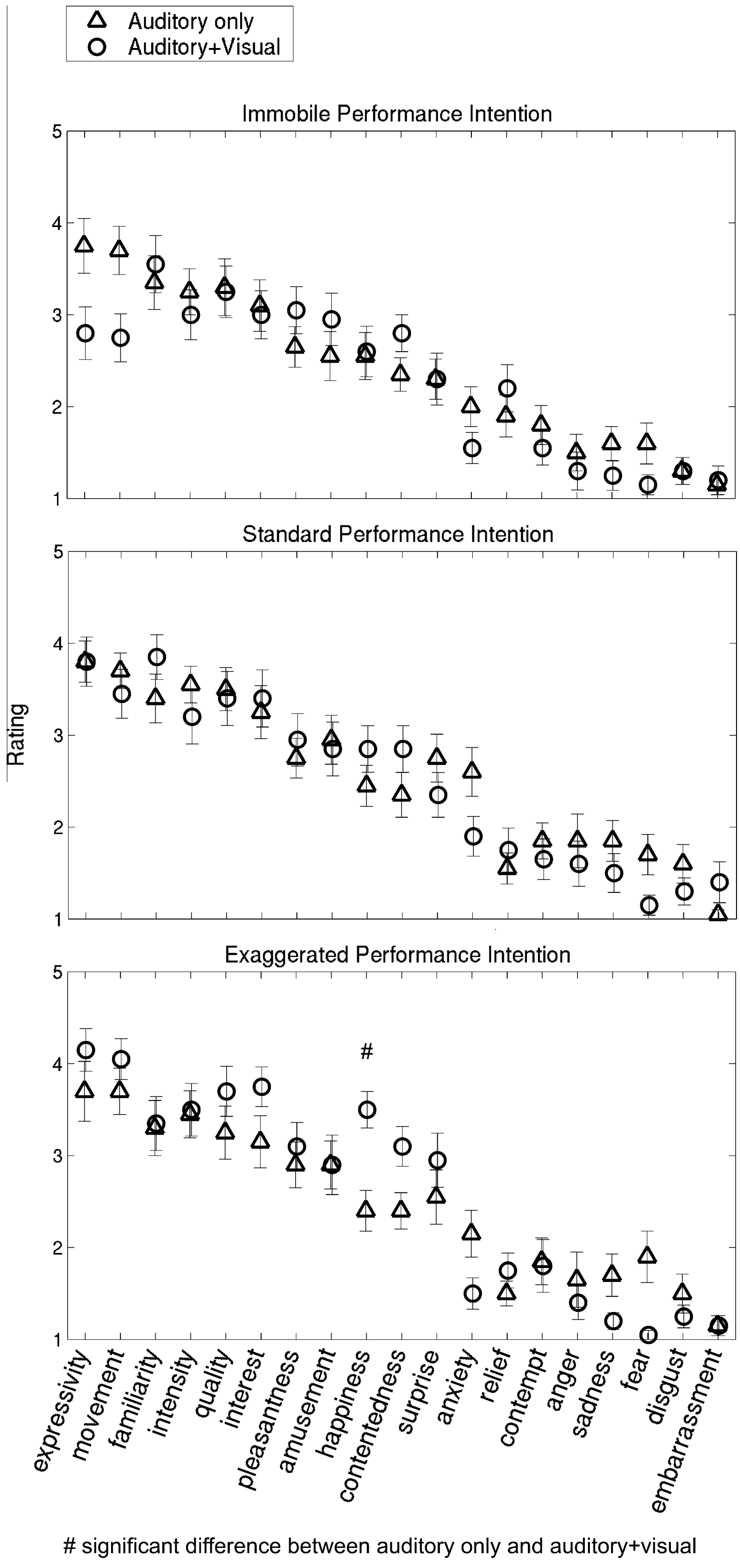

Fig. 3. This figure shows the AO and AV judgments of each word, for each performance manner. Symbols indicate significant differences, as noted at the bottom of the figure.

The factor analysis reduced the number of dependent variables from the full set of 19 words to a set of four uncorrelated factors. This four-factor solution converged in seven iterations. Fig. 4 shows the scree plot for the eigenvalue results (an index of the amount of variance accounted for by each factor). Fig. 5 depicts the evolution of the factors through the steps of the principal components analysis (Levitin, Schaaf, & Goldberg, 2005). The rotated factor loadings appear in Table 1. The four-factor solution accounted for 62% of the variance, which was an improvement of 6% over the three-factor solution.

We labeled each of the four factors as shown in Table 2. The name given to each factor was chosen by the authors to capture the semantic category of the clustered words, which appear to the right of the table. Researchers have made a distinction between words with a strong valence (e.g., disgust), and words with a weak valence (e.g., sadness), hence our decision to include the fourth factor as a separate grouping from the second factor (Green & Salovey, 1999; Tellegen, Watson, & Clark, 1999).

These results provide evidence that music is like complex social situations in which positive and negative emotions may or may not co-occur. There were independent dimensions for positive and negative valence, as well as a distinction between passive and active arousal (as in Cacioppo & Berntson, 1994; Cacioppo et al., 1997; Larsen

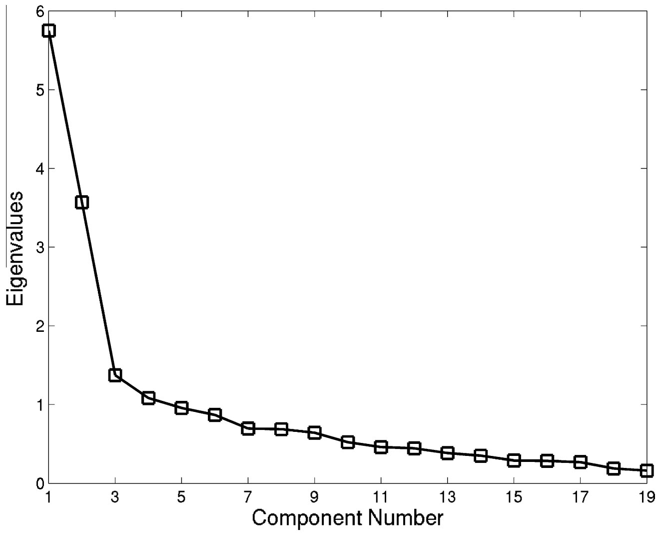

Fig. 4. Eigenvalues for the factor analysis results.

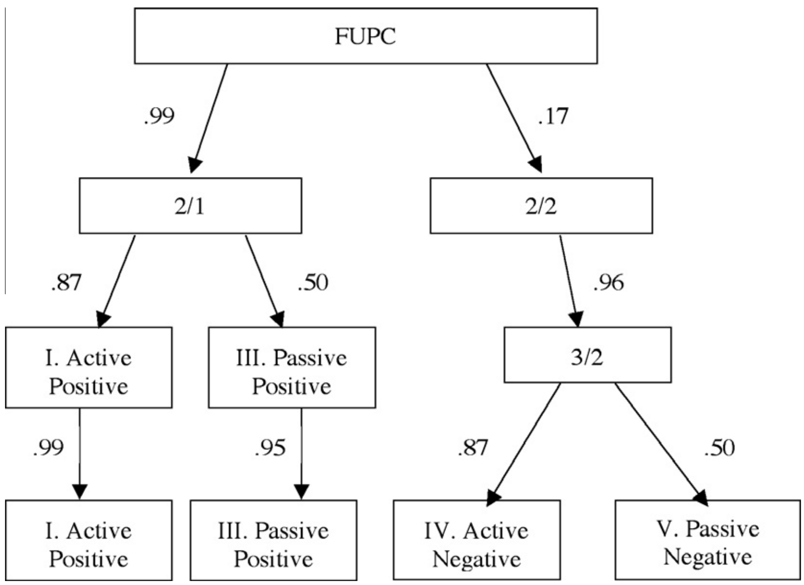

Fig. 5. A tree diagram depicting the hierarchical structure for the four extracted principal components. Successive levels reveal how the words became divided for each additional component. "FUPC" is the first unrotated principal component. Correlation values comparing factor scores from parent and offspring factors are shown next to the arrows that connect them. The initial split into two factors separated the negative from the positive words.

Table 1
Rotated component matrix.

	Component #			
	1	2	3	4
Expressivity	.86	.05	.17	.09
Intensity	.81	.19	.05	.21
Movement	.81	.16	.11	.06
Quality	.75	.03	.24	.13
Surprise	.70	.17	.07	−.08
Interest	.68	−.05	.42	−.05
Amusement	.59	−.21	.33	−.09
Disgust	−.10	.76	−.06	.22
Anxiety	.19	.76	−.11	.16
Anger	.13	.75	−.12	.17
Contempt	−.03	.75	.23	−.14
Fear	.18	.66	−.23	.25
Contentedness	.28	−.08	.77	−.15
Pleasantness	.43	−.11	.67	−.09
Relief	.07	.08	.62	.24
Happiness	.55	−.07	.61	−.12
Familiarity	.13	−.15	.56	.25
Embarrassment	−.05	.19	.05	.80
Sadness	.15	.24	.07	.66

et al., 2001; Tellegen et al., 1999; Watson & Tellegen, 1985).[1] The results of the factor analysis fit with the evaluative-space model, in which positive and negative emotions change independently.

One issue that is important in interpreting these results concerns the semantic distribution of the words. The orientation of affective dimensions might be skewed if a particular area of affective space is over-represented. In the rotated factor solution for the present set of independent variables, there were more Active Positive words (Factor #1) than Active Negative words (Factor #2), and more Passive Positive words (Factor #3) than Passive Negative words (Factor #4). Positive and negative valences were not equally represented, which might have affected the results. In order to address this issue, we reran the factor analysis in two ways: (1) using the two words with the highest correlation for each factor from the original analysis (expressivity, intensity, disgust, anxiety, contentedness, pleasantness, embarrassment, and sadness) and (2) using the two words with the lowest correlation for each factor from the original analysis (interest, amusement, contempt, fear, happiness, familiarity, embarrassment, sadness). For both analyses, the positive and negative valence words separated into separate orthogonal dimensions, in accordance with the analysis for the full set of 19 words. It is also notable that due to the relatively large number of variables (words) relative to the number of participant responses in this study, we must view the results of the factor analysis as a preliminary finding.

[1] To ensure that the orthogonal relationship between positive and negative affect was not simply an artifact of the varimax rotation, we calculated the unrotated solution scores. Even in the unrotated solution, the terms with a negative valence separated from those with a positive valence to form a separate and orthogonal dimension.

3.3. ANOVA of the factor scores

As part of the factor analysis, we recorded the factor scores for each participant on each orthogonal dimension. A factor score represents a participant's standard score on a particular dimension. We applied repeated-measures ANOVAs to the participants' factor scores, with a separate analysis for each of the four orthogonal dimensions. These ANOVAs included a between-subjects factor "mode of presentation" (AO, VO, AV), and within-subjects factors "performer" (performer W, performer R), and "manner" (restrained, standard, exaggerated).

3.4. Active positive dimension

Fig. 6 displays the mean active-positive factor scores along with error bars for the standard error of the mean. The repeated-measures ANOVA yielded significant main and interaction effects. There was a main effect of mode of presentation ($F(2, 27) = 3.49$; $p < .05$). A Tukey HSD post hoc analysis revealed a significant difference between the AO and VO modes of presentation overall.

There was also a significant main effect for performance manner ($F(2, 27) = 22.94$, $p < .001$). Combined Helmert and difference contrasts revealed that the mean for the restrained manner was significantly lower than both the standard and exaggerated manners. The results for the standard and exaggerated manners did not differ.

A significant interaction effect between performance manner and mode of presentation emerged as well ($F(4, 27) = 4.78$, $p = .002$). To determine the source of the interaction, we ran two sets of post hoc analyses. The first set of analyses considered whether the effect of presentation condition might have differed across performance manners. We ran three ANOVA's – one for each of the three performance manners. The between-subjects factor was "mode of presentation" (AO, VO, AV), and the within-subjects factor was "performer" (performer W, performer R). Only for the restrained performance manner was there a significant main effect of mode of presentation ($F(2, 27) = 8.88$, $p < .005$). Further post hoc analyses with a Holm–Bonferroni correction with a significance level of .05 revealed that ratings for the VO condition were significantly lower than those for both the AO and the AV conditions for the restrained performance manner. There were no differences across presentation conditions for the standard and exaggerated manners, which provides evidence that only watching the standard and exaggerated performances elicited an equivalent intensity of active-positive emotion as did only hearing, or both watching and hearing. Low ratings from the VO group in response to the restrained performances caused the main effect of mode of presentation, which we mentioned above.

The second set of post hoc analyses considered whether the effect of performance manner might have differed across presentation conditions. We ran three post hoc ANOVA's – one for each of the three presentation conditions. The within-subjects factors were "performer" (performer W, performer R), and "manner" (restrained, standard, exaggerated). In the VO condition, there was a significant

Table 2

Factor # and name	Variance accounted for (%)	Emotion terms
I. Active positive	24	Expressivity (.86), intensity (.81), movement (.81), quality (.75), surprise (70), interest (.68), amusement (.59)
II. Active negative	16	Disgust (.76), anxiety (.76), anger (.75), contempt (.75), fear (.66)
III. Passive positive	14	Contentedness (.77), pleasantness (.67), relief (.62), happiness (.61), familiarity (.56)
IV. Passive negative	8	Embarrassment (.80), sadness (.66)

Note: Parentheses contain the correlations between data for the original words and the corresponding scores on the extracted dimensions.

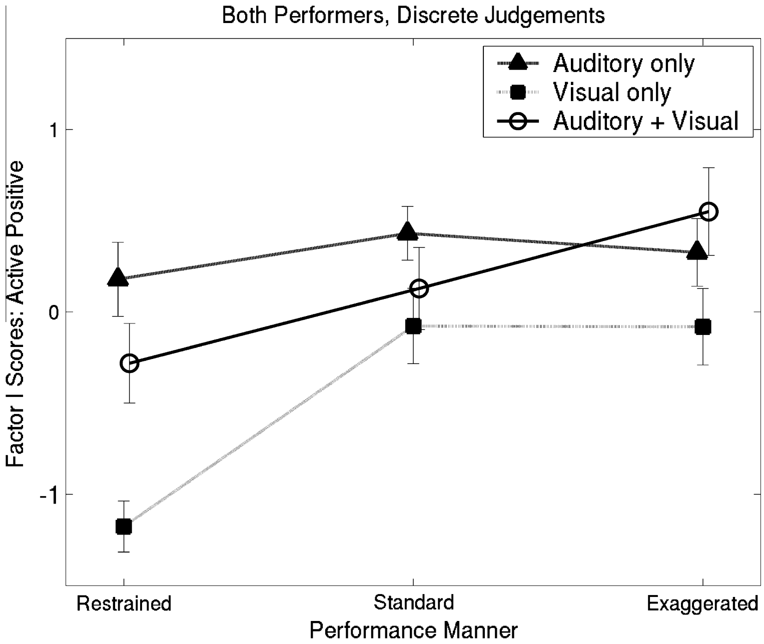

Fig. 6. Mean factor scores for the active positive dimension. Error bars represent one standard error of the mean.

main effect of performance manner ($F(2, 18) = 24.18$, $p < .001$); further post hoc analyses with a Holm–Bonferroni correction and a significance level of .05 revealed that ratings for the restrained manner were significantly lower than those for both the standard and exaggerated manners. There were no significant main effects in the AO condition, which provided evidence that the differences in performance manner had no effect on active-positive emotion for participants who only heard the music. In the AV condition, there was a significant main effect of performance manner ($F(2, 18) = 10.54$, $p < .001$). Further post hoc analyses with a Holm–Bonferroni correction and a significance level of .05 revealed that ratings for the exaggerated manner were significantly higher than those for both the standard and restrained manners. Notably, only for participants who could both hear and see the performances did the exaggerated performance manner lead to a significant increase in active-positive emotion compared to the standard performance manner. This finding provides evidence for an emergent effect in the AV condition. Vines and colleagues (2005, 2006) also reported an emergent intensity of emotional response for the AV mode of presentation.

3.5. Active negative dimension

There were no significant main or interaction effects for the active-negative factor scores. It appears that variations in performance manner, sensory modality of presentation, and performer do not modulate the intensity of active-negative emotion. These data suggest that the musical piece itself, as composed and dictated in the score, largely or entirely determines whether a performance will elicit active-negative emotions. This conclusion is not entirely intuitive, particularly with regard to the VO condition. However, previous research has shown that musicians effectively communicate their intended emotions through their body movements (Dahl & Friberg, 2007). Given that a musician's intended emotion will tend to align with the emotion of the musical piece, the composition will largely determine even the emotions that participants in the VO condition experience.

3.6. Passive-positive dimension

The repeated-measures ANOVA of the passive-positive scores revealed no main effects. However, there were significant interaction effects between the factors performer and manner ($F(2, 54) = 4.52$, $p = .015$), and among performer, manner, and mode of presentation ($F(4, 54) = 2.57$, $p = .048$). Fig. 7 shows the passive-positive scores, with the data displayed separately for each performer.

To examine the interaction between the factors performer and manner, we ran two post hoc one-way repeated-measures ANOVA's (one for each performer) to compare the data for the three performance manners (restrained, standard, and exaggerated). For performer R, there was no significant difference across performance manners ($F(2, 58) = .64$, $p = .53$). However, there was a significant effect of performance manner for performer W ($F(2, 58) = 5.77$, $p = .005$). Pairwise comparisons with a Holm–Bonferroni correction at the .05 significance level revealed that the passive-positive data for the restrained manner were significantly higher compared to both the data for the standard and exaggerated manners. Notably, performer W's restrained performance elicited a greater degree of passive-positive emotion compared to either his standard or exaggerated performances. This pattern was unique to performer W.

To examine the interaction for the factors performer, manner, and mode of presentation, we ran three post hoc one-way repeated-measures ANOVA's (one for each mode of presentation) on the data for performer W. These ANOVA's compared the three performance manners (restrained, standard, and exaggerated). For the AO condition, there was no significant difference across performance manners ($F(2, 18) = .90$, $p = .43$). However, there was a significant effect of performance manner in the VO condition ($F(2, 18) = 6.1$, $p < .01$). Pairwise comparisons with a Holm–Bonferroni correction at the .05 significance level revealed that data for the restrained manner were significantly higher than for the standard manner. The test for the AV condition fell short of significance ($F(2, 18) = 2.9$, $p = .084$), though the trend was similar to that for the VO condition.

These results showed that the trend for performer W was driven by ratings from participants who could see the performances. This provided evidence that idiosyncratic characteristics of the performers' movements (or of the individual performances) had differential effects on passive-positive emotion. For performer W, his natural movements in the standard and exaggerated conditions actually dampened the experience of passive-positive emotion for observers. We posit that in some circumstances, a musician may actually increase certain aspects of positive emotional response in observers by consciously inhibiting body movement.

3.7. Passive negative dimension

The repeated-measures ANOVA revealed neither significant main nor interaction effects. Therefore, as with the active negative dimension, we conclude that variations in mode of presentation, performer, and performance manner do not have differential effects on passive-negative emotions.

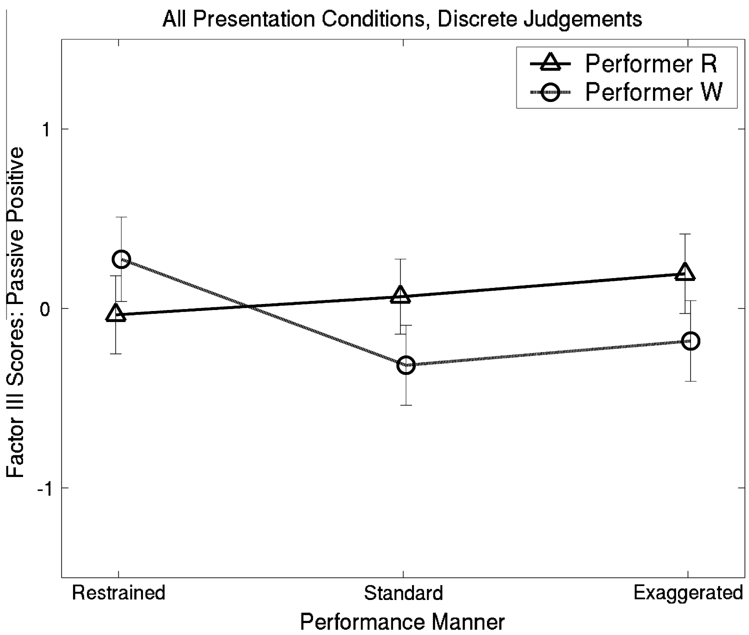

Fig. 7. Visualizing the interaction between the factors performer and manner. Factor scores for the passive-positive dimension are plotted against performance manner. Error bars represent one standard error of the mean.

4. Conclusions

We found that whether a subject could see a performance was the most important factor in determining how musicians' expressive intentions would affect the emotions conveyed. The performers' intended level of expressivity did not affect the emotions conveyed by sound alone. In contrast, participants who could see the performances made ratings that distinguished between the expressive levels for several of the words. This was the case for the active positive dimension of the factor analysis as well, as shown in Fig. 6.

This study provides evidence for an emergent intensity of positive emotion when musical performance is both seen and heard. For the emotion *happiness*, AV participants' judgments were significantly higher than those for the AO participants, in response to the exaggerated performance manner. Therefore, being able to see in addition to hear the performances led to a more intense experience of happiness. The analysis of the factor scores revealed an emergent effect as well. Only AV participants experienced an increase in active-positive emotion for the exaggerated manner compared to the standard manner of performance. Taken together, these results suggest that seeing a musician perform may increase the potential for observers to experience positive emotion.

Notably, the performers' movements did not always lead to an increase in positive emotion. For one of the two clarinetists, restraining body movements actually led to an increase in passive-positive emotion. It appears that there is not a linear relationship between the amount of body movement and the intensity of positive emotion that a performance conveys. Instead, the effect of body movement may depend upon idiosyncratic characteristics of each performer, or of each performance.

We also found evidence that music has the potential to generate multi-valent emotional states in which positive and negative feelings co-occur, much like life experiences with complex antecedents and consequences. A factor analysis revealed that the musical performances induced an experience involving orthogonal dimensions for positive and negative emotion. The evaluative-space theory (Cacioppo & Berntson, 1994; Cacioppo et al., 1997) provided the best fit for these data. Whether this finding is indicative of music in general, or is limited to the particular genre studied here is a matter for future research.

Our findings support the theoretical perspective that the visual component of musical performance makes a unique contribution to the communication of emotion from performer to audience. The results of this study are in accordance with previous research showing that speakers' facial expressions and gestures carry information that is not available in aural speech alone (Goldin-Meadow, 2003; McNeill, 1992, 1999, 2005), and that musical emotions are communicated by musicians' movements in addition to the sound (Di Carlo & Guaitella, 2004). It is notable that studies have found that visual information is particularly valuable in speech perception when there is some ambiguity in the sound, for example when the speech is embedded in noise (Schwartz, Berthommier, & Savariaux, 2004). This may be the case in the context of musical performance as well. Stravinsky's Second Piece for Clarinet Solo has an ambiguous tonal structure. The musicians' movements may offer cues that help an observer to resolve the ambiguity in sound by providing further information about the emotional content of the piece. Based upon this idea, we hypothesize that the more unfamiliar observers are with the music, the more they will rely on visual cues to determine the emotional content of the work. Along these lines, Davidson and Correia (2002) noted that audience members who are not skilled at listening rely entirely on visual cues to determine the emotional content of a musical piece.

We have found strong evidence that the visual component of musical performance makes a unique contribution to the communication of emotion from performer to audience. Seeing a musician can augment, complement, and interact with the sound to modify the overall experience of music. The emotions in music, and the performers' intended expressive qualities are conveyed in differing ways (and to differing degrees) by sound, sight, and the combination of senses. This study shows that the communication and perception of musical emotions depend upon interactions between sensory modalities, as well as unique channels of visual and auditory information. The emotions that a performance does convey appear to be structured like complex life emotions.

Acknowledgements

The research reported herein was submitted in partial fulfillment for the Ph.D. Degree at McGill University by the first author, under the supervision of the last author. We thank the performers, and Nora Hussein, Hadiya Nedd-Roderique, and Sawsan M'Birkou for their help with the data collection and video analysis.

This research was supported by grants to DJL from NSERC (Natural Science and Engineering Research Council of Canada), John & Ethelene Gareau, and Google; FQRNT Strategic Professor awards to DJL and MMW; a Guggenheim Fellowship and NSF (National Science Foundation) grant to CLK; and grants to BWV from the J. W. McConnell Foundation, CIRMMT, and the Michael Smith Foundation for Health Research.

References

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.

Bharucha, J. J., Curtis, M., & Paroo, K. (2006). Varieties of musical experience. *Cognition, 100*, 131–172.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion, 19*(8), 1113–1139.

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in the brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences, 98*, 11810–11823.

Broughton, M., Stevens, C., & Malloch, S. (2009). Music, movement and marimba: An investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychology of Music, 37*(2), 137–153.

Brown, S. (2000). The 'musilanguage' model of music evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 271–300). Cambridge: MIT Press.
Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115*, 401–423.
Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review, 1*, 3–25.
Chapados, C., & Levitin, D. J. (2008). Cross-modal interactions in the experience of musical performances: Physiological correlates. *Cognition, 108*, 639–651.
Cooper, G. W., & Meyer, L. B. (1960). *The rhythmic structure of music*. Chicago: University of Chicago Press.
Cross, I. (2003). Music as a biocultural phenomenon. *Annals of the New York Academy of Sciences, 999*, 106–111.
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.
Davidson, J. W. (1994). Which areas of a pianist's body convey information about expressive intention to an audience? *Journal of Human Movement Studies, 26*, 279–301.
Davidson, J. W., & Correia, J. S. (2002). Body movement. In R. Parncutt & G. E. McPherson (Eds.), *The science and psychology of music performance* (pp. 237–253). New York: Oxford University Press.
Davies, S. (1994). *Musical meaning and expression*. Ithaca: Cornell University Press.
Delalande, F. (1988). La gestique de Gould; éléments pour une sémiologie du geste musical. In G. Guertin (Ed.), *Glenn Gould Pluriel* (pp. 83–111). Montréal: Louise Courteau Éditrice Inc..
Deliège, I., & El Ahmade, A. (1990). Mechanisms of cue extraction in musical groupings: A study of perception on *Sequenza VI* for viola solo by Luciano Berio. *Psychology of Music, 18*, 18–44.
Di Carlo, N. S., & Guaitella, I. (2004). Facial expressions of emotion in speech and singing. *Semiotica, 149*(1/4), 37–55.
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*, 727–738.
Eibl-Eibesfeldt, I. (1972). Similarities and differences between cultures in expressive movements. In R. A. Hinde (Ed.), *Non-verbal communication* (pp. 297–314). Cambridge: Cambridge University Press.
Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*, 169–200.
Ekman, P. (1998). Facial expressions. In T. Dalgleish & T. Power (Eds.), *The handbook of cognition and emotion* (pp. 301–320). Sussex, UK: John Wiley & Sons, Ltd..
Faith, M., & Thayer, J. F. (2001). A dynamical systems interpretation of a dimensional model of emotion. *Scandinavian Journal of Psychology, 42*, 121–133.
Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition, 100*(1), 173–215.
Fredrickson, W. E. (2000). Perception of tension in music: Musicians versus nonmusicians. *Journal of Music Therapy, 37*(1), 40–50.
Gabrielsson, A. (1999). The performance of music. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 501–602). London: Academic Press.
Gabrielsson, A. (2001). Emotions in strong experiences with music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 431–449). Oxford: Oxford University Press.
Gabrielsson, A., & Lindström, S. (2003). Strong experiences related to music: A descriptive system. *Musicae Scientiae, 7*(2), 157–217.
Galati, G., Committeri, G., Spitoni, G., Aprile, T., Di Russo, F., Pitzalis, S., et al. (2008). A selective representation of the meaning of actions in the auditory mirror system. *Neuroimage, 40*(3), 1274–1286.
Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology, 16*(18), 1824–1829.
Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature, 462*, 502–504.
Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.
Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology, 64*(6), 1029–1041.
Green, D. P., & Salovey, P. (1999). In what sense are positive and negative affect independent? A reply to Tellegen, Watson, and Clark. *Psychological Science, 10*(4), 304–306.
Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion*. New York: Cambridge University Press.
Hauk, O., Shtyrov, Y., & Pulvermuller, F. (2006). The sound of actions as reflected by mismatch negativity: Rapid activation of cortical sensory–motor networks by sounds associated with finger and tongue movements. *European Journal of Neuroscience, 23*(3), 811–821.
Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition & Emotion, 22*(2), 327–352.
Huron, D. (2001). Is music evolutionary adaptation? *Annals of the New York Academy of Sciences, 930*, 43–61.
Izard, C. E. (1971). *The face of emotion*. New York: Appleton Century Crofts.
Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition, 100*, 33–72.
Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*, 201–211.
Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience, 7*(3), 302–307.
Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.
Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception, 13*(3), 401–432.
Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology, 51*(4), 336–352.
Krumhansl, C. L. (2002). Music: A link between cognition and emotion. *Current Directions in Psychological Science, 11*(2), 45–50.
Krumhansl, C. L., & Schenck, D. L. (1997). Can dance reflect the structural and expressive qualities of music? A perceptual experiment on Balanchine's choreography of Mozart's Di-vertimento No. 15. *Musicae Scientiae, 1*, 63–85.
Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2004). The evaluative space grid: A single-item measure of positive and negative evaluative reactions. Working paper, Department of Psychology, Texas Tech University, Lubbock.
Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology, 81*, 684–696.
Lasswell, H. D. (1964). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas*. New York: Cooper Square Publishers.
Lerdahl, F. (2001). *Tonal pitch space*. New York: Oxford University Press.
Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
Levinson, J. (1980). What a musical work is. *Journal of Philosophy, 77*, 5–28.
Levitin, D. J., & Menon, V. (2003). Musical structure is processed in "language" areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *Neuroimage, 20*(4), 2142–2152.
Levitin, D. J., & Menon, V. (2005). The neural locus of temporal structure and expectancies in music: Evidence from functional neuroimaging at 3 Tesla. *Music Perception, 22*(3), 549–562.
Levitin, D. J., Schaaf, A., & Goldberg, L. R. (2005). *Factor diagrammer (Version 1.1b)*. Montreal: McGill University (computer software) <http://www.psych.mcgill.ca/labs/levitin/software/factor_diagrammer/>.
Lopes, D. M., & Bergeron, V. (2009). Hearing and seeing musical expression. *Philosophy and Phenomenological Research, 78*, 1–16.
McAdams, S., Vines, B. W., Vieillard, S., Smith, B. K., & Reynolds, R. (2004). Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting. *Music Perception, 22*(2), 297–350.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748.
McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
McNeill, D. (1999). Triangulating the growth point—arriving at consciousness. In L. S. Messing & R. Campbell (Eds.), *Gesture, speech, and sign* (pp. 77–92). New York: Oxford University Press.
McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
Menon, V., & Levitin, D. J. (2005). The rewards of music listening: Response and physiological connectivity of the mesolimbic system. *Neuroimage, 28*(1), 175–184.
Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.
Mithen, S. (2007). *The singing Neanderthals*. Cambridge, MA: Harvard University Press.
Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review, 97*(3), 315–331.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674–681.
Repp, B. H. (1996). Patterns of note onset asynchronies in expressive piano performance. *Journal of the Acoustical Society of America, 100*, 3917–3932.
Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectations, gender, recognition, and deceptive intention. *Journal of Experimental Psychology: General, 112*, 585–615.
Russell, J. A. (1979). Affective space is bipolar. *Journal of Personality and Social Psychology, 37*, 345–356.
Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161–1178.
Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145–172.
Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin, 125*, 3–30.
Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two dimensional emotion space. *Australian Journal of Psychology, 51*, 154–165.
Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception, 21*(4), 561–585.
Schutz, M. (2008). Seeing music? What musicians need to know about vision. *Empirical Musicology Review, 3*(3), 83–108.
Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception, 36*(6), 888–897.
Schwartz, J.-L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio–visual interactions in speech identification. *Cognition, 93*(2), B69–B78.
Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science, 10*(4, Special Section), 297–303.
Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica, 156*, 203–227.
Thompson, W. F., Russo, F., & Quinto, L. (2008). Audio–visual integration of emotional cues in song. *Cognition & Emotion, 22*(8), 1457–1470.
Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition, 101*(1), 80–103.
Vines, B. W., Nuzzo, R. L., & Levitin, D. J. (2005). Quantifying and analyzing musical dynamics: Differential calculus, physics and functional data techniques. *Music Perception, 23*(2), 137–152.
Wanderley, M. M. (2002). Quantitative analysis of non-obvious performer gestures. In I. Wachsmuth & T. Sowa (Eds.), *Gesture and sign language in human–computer interaction* (pp. 241–253). Berlin: Springer Verlag.
Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*(2), 219–235.