Seeing music performance: Visual influences on perception and experience

WILLIAM FORDE THOMPSON, PHIL GRAHAM, and FRANK A. RUSSO

Abstract

Drawing from ethnographic, empirical, and historical / cultural perspectives, we examine the extent to which visual aspects of music contribute to the communication that takes place between performers and their listeners. First, we introduce a framework for understanding how media and genres shape aural and visual experiences of music. Second, we present case studies of two performances, and describe the relation between visual and aural aspects of performance. Third, we report empirical evidence that visual aspects of performance reliably influence perceptions of musical structure (pitch related features) and affective interpretations of music. Finally, we trace new and old media trajectories of aural and visual dimensions of music, and highlight how our conceptions, perceptions and appreciation of music are intertwined with technological innovation and media deployment strategies.

1. Introduction

Making music involves not only the communication of musical sounds but is also characterized by a continuously changing and meaningful use of facial expressions, body movements, and hand gestures. Until the late nineteenth century, music performances were almost always experienced as audio-visually integrated activities. Audio and visual components of music performance were separated with the introduction of technologies such as the radio and gramophone, which isolated the aural component of music from all other aspects. That separation has influenced conceptions of music, such that visual contributions to music activities are often ignored.

Facial expressions and gestures affect music experience on several levels. At a basic level, visual information often signals the timing of

Semiotica 156–1/4 (2005), 177–201 0037–1998/05/0156–0177
© Walter de Gruyter

(V7(M) 10/5/05 15:55) WDG/G J-1299 Semiotica, 156 PMU: WSL(V) 10/05/05 Times_M (0).3.04.05 (148×225mm) pp. 177–202 007_P (p. 177)

musical events, focusing listeners' attention to (or away from) critical acoustic information at specific moments in time. By directing attention in this way, visual cues can increase or decrease musical intelligibility. To draw an analogy, it is well known that seeing the face of a speaker can increase the intelligibility of speech in a noisy environment. Similar effects are observed for musical intelligibility.

At a perceptual level, they signal important melodic, harmonic, and rhythmic events. Facial expressions may reflect the additional concentration that is needed to perform notes or passages that are unexpected or tonally unstable. Performers may also intentionally introduce facial expressions and other gestures as a way of sharing with listeners their understanding of the musical significance of such events. Facial expressions and gestures are also used to convey the performer's understanding of segmentation (points of closure), intervallic information (whether a melodic interval is large or small), and points of expectancy fulfillment or violation. In this way, visual aspects of performance signal that performers are not merely producers of sound but are themselves listeners, highlighting the musical activity as a shared experience between performers and listeners.

At the level of affect, music is deeply infused with human emotion and performers use a variety of resources to express and convey that emotional content. Emotions are communicated not only in sounded events but also in facial expressions and other bodily gestures (Di Carlo and Guaitella 2004). It has been shown that musically unschooled listeners may rely more heavily on visual than on aural cues when evaluating affective meaning in music (Davidson and Correia 2002).

Finally, visual information is highly effective at conveying persona and attitude, at building up what is described from a systemic functionalist perspective as the 'interpersonal' (Halliday 1994) or 'attitudinal' (Lemke 1995) dimension of meaning.[1] Facial expressions and hand gestures allow performers to *cozy up* to the audience, emphasizing the music performance as reciprocal human interaction, whereas an absence of visual information leaves an impression that the performance is a solitary act in which the listener's role is primarily that of a *voyeur*. That is, visual aspects of music personalize the music, drawing performers and listeners closer together in a shared experience.

In this paper, we consider ethnographic, empirical, and historical / cultural perspectives in order to examine the extent to which visual and aural aspects of music are integrated in musical experience. We first introduce a framework for understanding how different media (the technological means of distribution), genres (regularized patterns of expression), and modes of transmission (expressive resources such as facial

expressions, gestures, and melodies) combine to shape our conception and experience of music. We next present case studies of two performances, and describe the relation between visual and aural aspects of performance. We next report empirical findings that support a thesis of cognitively integrated rather than isolated visual and aural faculties. Our data indicate that visual aspects of performance reliably affect perceptions of musical structure (pitch related features) and affective interpretations of music. Finally, we discuss historic and social implications of music performance as a mediated or 'technologized' event: as artificially separating the aural from the visual in the case of radio and gramophone; or as an embodied, personal experience. By tracing new and old media trajectories of aural and visual dimensions of music, we highlight how conceptions, perceptions, and appreciation of music are intertwined with technological innovation and media deployment strategies. We conclude by arguing that visual aspects of music — hidden as they once were within the mediations permitted by radio and gramophone — have more recently become privileged media forms, resulting in the valorization of artists, music, and performance types in a predominantly visual media environment.

2. Modes, genres, and medium

Music is sometimes experienced as a live performance but is more often experienced through a variety of media technologies such as radio, television, film, iPod, or the internet. We first outline a framework for understanding the role of media and genres in enhancing, editing, or censoring visual and aural aspects of experience. The technological character of the medium means that a filmed music performance provides a different experience from the same performance broadcast on radio. Aural and visual modes of expression are also dependent on the musical genre, with some genres emphasizing visual modes of expression more than others. In short, music experience may be understood with reference to multiple levels of influence, with the technological character of the medium constraining the kind of genres and modes that can be transmitted, and musical genres differentially highlighting or filtering aural and visual modes of expression.

Figure 1 illustrates these levels of influence. A medium is defined as the channel through which communication takes place, whether writing, speech, television, or internet. Genres describe repetitions of patterned interactions within and across cultures. They are typical doings of a community that are repeatable and recognized as the same type from one occurrence to another: a blues performance, baseball game, train ride,

(V7(M) 10/5/05 15:56) WDG/G J-1299 Semiotica, 156 PMU: WSL(V) 10/05/05 Times_M (0).3.04.05 (148×225mm) pp. 177–202 007_P (p. 179)

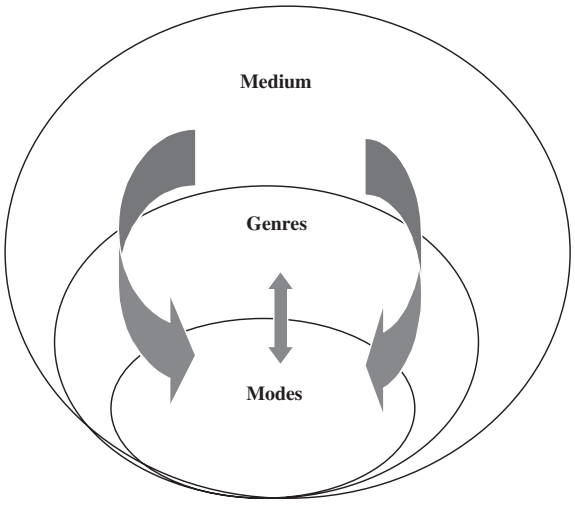

Figure 1. *Three levels in which performances may be interpreted and experienced*

writing a check, making a phone call (Lemke 1998). Although a given medium will accommodate a wide range of genres, the technological properties of a medium may be better suited to some genres than others. Each genre, in turn, is constituted by multiple modes, such as tone of voice, words, gestures, and facial expressions. Modes are the resources of expression with which genres are textured, constituted, or formed (Graham 2003).

A genre is never independent of technologies or mediation processes, so any account of genre must include an account of its technological aspects and the modes that constitute a given genre. Media exert their most apparent constraints on the modes they accommodate, selecting filtering or emphasizing them: one can neither show facial expressions through radio nor timbre of voice through text. In film, camera shots and post-production techniques such as automatic dialogue replacement (ADR) draw attention to certain modes at the expense of others. Medium works in a 'downwards' way upon genre formations, constraining the range of constituency elements that comprise a given genre by limiting the range of modes by which meanings are made and conveyed.

3. A case study of two performances

We now summarize the use of facial expression, body movement, and gesture in two filmed performances, one by B. B. King and one by Judy

Garland. Descriptions are drawn from structured interviews by a trained musicologist (Jeff Cupchik) of two other musicologists who observed and commented on the performance by B. B. King (Judges A and B), and from our observations of the performance by Judy Garland. Our aim was to identify intentional aesthetic movements and gestures that serve to highlight, articulate, interpret, and clarify the music; that act to communicate emotion or personality; and that otherwise elicit specific interactions between performers and listeners. Whereas certain body movements are required in order to sing or play an instrument, others may function to encourage listeners to attend to certain dimensions of the music rather than others, to interpret those dimensions in specific ways, and to experience the event as a social interaction between performers and listeners.

According to Kurosawa and Davidson (2005), the facial expressions and gestures used in music performance can be interpreted in view of the categories of non-verbal behaviour described by Ekman and Friesen (1981), which include: *emblems*, *illustrators*, *regulators*, and *affect displays. Emblems* are body movements with a meaning that is shared by members of a group, class, or culture, and that can be translated into a verbal message. They include the thumbs up sign, positioning fingers in a peace sign, and 'sticking out' the tongue. *Illustrators* are used to clarify or emphasize the content of a message. They include gestures that identify people or objects (e.g., by pointing), accentuate words or phrases, represent relations between objects or ideas, or illustrate the timing of an event. *Regulators* are gestures that maintain the pace and content of interactions, and include head nods and eye contact. Finally, *affect displays* are expressions that indicate emotional states, such as smiling or frowning.

In the instrumental performance by B. B. King (*Blues Boys Tune*), facial expressions often function both as *affect displays* and as musical *illustrators* to emphasize occurrences of dissonance and 'blue' notes (Pearce 2003). King frequently adopts an introspective demeanor, with eyes closed and a pained expression, yet stubbornly shaking his head. This *affective display* conveys an impression of stoically reflecting upon but not surrendering to difficult emotions. Periodically he stares open-eyed at the audience with an open mouth. The expression appears to convey a sense of wonder but was also described by one of our judges as an expression of 'checking in' with the audience (i.e., functioning as a *regulator* in the performer-audience interaction). There are also structural movements that serve as musical illustrators: he suddenly looks up when the pitch bends upward, or arches his back when there is a prolonged point of energy (see figure 2).

(V7(M) 10/5/05 15:56) WDG/G J-1299 Semiotica, 156 PMU: WSL(V) 10/05/05 Times_M (0).3.04.05 (148×225mm) pp. 177–202 007_P (p. 181)

Figure 2. *B. B. King in performance*

It is notable that B. B. King's facial expressions closely track his guitar sounds and not the accompanying instruments, such as percussion or other guitars. In some cases his rapid head shaking movement mirrors vibrato on individual notes. This gesture has the effect of drawing the listeners' attention to local aspects of music, specifically to B. B. King's nuanced treatment of individual notes, and away from larger-scale musical structure. But his body movements also reflect large-scale structure, becoming generally more pronounced as the performance moves to the point of climax.

Judge A observed that King's facial expressions often functioned to signal that certain passages were difficult but satisfying to play. He noted that the expressions closely parallel the call and response structure of blues music, even though the audience in this filmed performance was disengaged from any invitation to interact in this way. He also suggested that King's open-mouth expression conveys a sense of letting the sound go free.

It looks like he's working on every detail when he's got his mouth closed. When he opens his mouth it has to do with letting the sound sing. It's as though he's saying 'What do you think about this?' His face conveys the chronological process

of getting a musical idea and realizing it in sound. When he's got the corner of his mouth lifted he's *thinking* of an idea; when he is wincing he's doing the *technical* work. And when he's satisfied with it and wants the audience to respond, he opens his mouth and leans forward. People react when he leans forward and opens his mouth and eyes wide, because they know he's asking for a reaction.

Judge B pointed out that King's movements help him to get into the feel of the music.

I don't know where it comes from, but as a guitarist you just do it. Jimi Hendrix, Eric Clapton, all these guys, they bend notes and they lean back. It's a physical thing. Like you get leverage. It feels like you get leverage when you do that.

Judy Garland is well known as a vocalist for having developed a highly stylized and complex visual style that functions on many levels. In the performance examined (*Just in time*), gestures and facial expressions are closely tied to the narrative of the lyric, frequently functioning as *illustrators* of verbal content (Jewison 1962). For example, the word 'tossed' is accompanied by a 'tossing' gesture. At a point where the lyric conveys a negative emotion ('I was lost'), hand gestures simulate 'swimming' motions, completing the telling image of being lost at sea. At the same time, her warm smile provides an *affect display* that anticipates the happy resolution still to come ('now you're here'). By snapping fingers on the lyrics 'for love came just in time' she provides a rhythmic *illustrator* of the verbal content, as if the beat is a metaphor for love arriving 'just in time.' Near the end of the performance a winning fist thrust into the air is *emblematic* of the successful resolution to the story narrative, 'That lucky day.'

In addition, facial expressions and gestures also reflect elements of the music that cannot be predicted just from the lyrics. As an example, at a point of tonal modulation, Garland walks boldly forward to illustrat the significance of this musical change. Hand gestures often also mirror pitch height, as when a low note and the lyric 'low' are accompanied by an outstretched arm with her hand at a low spatial position. An upward hand movement also mirrors the rising pitch in the melodic line that accompanies the desperate words 'nowhere to go' (see Figure 3).

It is of technological significance, specifically cinematic in character, that in Judy Garland's live 1962 performance the camera focuses on her whole body, suggesting that she is a total performer. In contrast, perhaps in the wake of MTV, more contemporary and highly edited cinematography works montage-like with extreme close-up and fast-cuts to 'dismember' bodies, lingering on isolated body parts (Kilbourne 1999). Such techniques impact on the role of the visual in the delivery of music

(V7(M) 10/5/05 15:56) WDG/G J-1299 Semiotica, 156 PMU: WSL(V) 10/05/05 Times_M (0).3.04.05 (148×225mm) pp. 177–202 007_P (p. 183)

Figure 3. *Judy Garland in performance*

performance, permitting in the former case Garland's whole-of-person narrative, and in other cases providing visual enhancements to a performance perhaps in order to compensate for or distract from its lack of musicality.

4. Psychological implications

In this section we explore empirical research on visual aspects of music performance. Our aim in conducting empirical research was to determine whether visual aspects of performance actually influence perceptual and aesthetic judgements of music, or whether they are interpreted as playing merely a subsidiary or non-essential role. Psychological research on music performance has grown enormously in the past two decades, although very few researchers have examined the perceptual consequences of visual aspects of music performance. After reviewing research on music performance, Gabrielsson concluded that there is a 'need for investigating the role that visual information may play in music perception' (Gabrielsson 1999: 523). He emphasized that the conceptualization of music as a purely acoustic phenomenon is not typical of all cultures or all times.

Rather, music making is historically and typically experienced as bodily-mediated phenomena, as events in which people interact with each other *in person*.

A small number of researchers have pointed to the importance of visual aspects of performance. Juslin (2001) noted that highly emotional facial expressions often accompany music performances. Davidson (1993, 1994, 1995, 2001; Davidson and Correia 2002) provided detailed analyses of visual aspects of performances and reported that listeners are often influenced by such accompanying information. Clayton (forthcoming) described ethnographic research suggesting that gesture, motion, and facial expression can convey a performer's interpretation of music. Finally, Vines, Nuzzo and Levitin (in press) observed that the body movements of a performer mirror important structural elements in music.

Nonetheless, most psychological research on performance has ignored non-acoustic aspects of performance, considering them as extraneous and not essential to the music. Indeed, visual information related to music performance is often trivialized on the grounds that it is 'determined by the sound ... in order to produce a given chord, you have to place your fingers on given frets ...' (Cook 1998: 263). When visual influences on judgments of music are noted, they are often discussed as examples of visual 'bias' (McPherson and Thompson 1998; Thompson et al. 1998). For example, studies of performance adjudication indicate that skin color and gender influence performance assessments. For this reason orchestras and professional ensembles often have 'blind' auditions as the first stage for selecting members. Although bias effects are clearly not desirable, visual aspects of music performance can have aesthetic and perceptual consequences that positively contribute to and enhance the musical experience. In particular, facial expressions and bodily movements that occur during a music performance may greatly add to our experience of that music.

The ability to integrate information from different modalities in order to make sense of the world is well documented (Marks 1978; McGurk and MacDonald 1976). For example, visual information associated with speaking (i.e., facial movements) has surprisingly strong effects on speech perception. The McGurk effect (McGurk and MacDonald 1976) illustrates that visual articulatory information has striking effects on syllable perception. In some cases visual signals determine the syllable perceived; in other cases, acoustic and visual signals combine to produce a new perceived syllable. For example, if an acoustic signal created by uttering the syllable 'ba' is paired with facial movements associated with the syllable 'ga,' most people hear the 'compromise' syllable 'da.' If an acoustic signal for the syllable 'ba' is paired with facial movements associated

(V7(M) 10/5/05 15:56) WDG/G J-1299 Semiotica, 156 PMU: WSL(V) 10/05/05 Times_M (0).3.04.05 (148×225mm) pp. 177–202 007_P (p. 185)

with the syllable 'va,' most people hear the syllable 'va,' with the visual information overriding the acoustic signal. Thus, visual information is closely integrated with acoustic information in speech perception. At the syllabic level, perception is influenced by the relative strengths of acoustic and visual information, and whether or not a sensible compromise can be achieved. Integration occurs even when people are told of the dubbing procedure, suggesting that the process is automatic and unconscious. Again, these data foreground the potential difficulty associated with confining research in music to 'acoustic-only' information. They also foreground the strong influence that audio technologies have had in the last century on psychological research into music.

Theoretical discussions of connections between visual and musical information are not new. Dissanayake argues that music actually evolved in humans as 'multimedially presented and multimodally processed activity of temporally and spatially patterned vocal, bodily, and facial movements' (Dissanayake 2001: 389). Cook (1998) provides a theoretical discussion of multimodal music performance, such as ballet, opera, and MTV clips. Empirical studies of multimedia suggest that visual and musical information combine to form an integrated impression of an event. Such effects are well documented in studies of music for film and television (Cohen 2001; Gorbman 1987; Thompson 2002; Thompson et al. 1994). However, the latter studies concern the effects of music (underscoring) on judgments of film (primarily the visual component) rather than the reverse: the effects of visual information on judgments of music — an important distinction in the present discussion. Moreover, research on film music usually concerns the effects of combining music with visual images that do not correspond to bodily performances of that music. Soundtracks function as emotional or dramatic wallpaper for the visually dominated narrative.

However, few psychological researchers have identified 'musical McGurk effects' in which visual aspects of a music performance influence perceptions of musical structure itself. Saldaña and Rosenblum (1993) have identified the effects of visual information on the perception of plucked versus bowed sounds from string players. That is, given an audio-visual presentation of a performed note, a note may be 'heard' as starting somewhat more abruptly if the visual information suggests a plucked action than if it suggests a bowed action. The complex and highly expressive visual information conveyed through facial expressions and body movements in music performance suggests a wide range of potential effects that still need to be examined in depth. In the following section, we describe five laboratory experiments designed to contribute to such a research agenda.

The experiments described below were designed to identify visual influences on two types of musical judgments. In three experiments we assessed visual influences on structural interpretations of music and in two experiments we assessed visual influences on emotional interpretations of music. In each experiment, we presented listeners with video-clips of excerpts from performances by skilled musicians, either created by us or taken from archival sources. For each experiment we confirmed the reliability of our results using analysis of variance and an alpha level of 0.05.

4.1. *Effects of facial expression on the perception of music structure*

Musical dissonance occurs when a pitch or pitch combination does not fit with the overall harmonic or tonal context of a piece of music; consonance is to some degree a function of genre and refers to a note or notes fulfilling the expectations of harmonic and tonal context.[2] Sensitivity to relations between different notes, also called *relative pitch*, is at the foundation of our perception of melodic structure. Two sequential notes may be perceived as very different, forming a large melodic interval, or similar, forming a small interval. Certain pitch relations have special significance in music, such as those associated with third, fifth, and octave notes in a scale.

Facial expressions and gestures during performance may influence perceptions of consonance or pitch because they reflect the physical effort required to perform a musical interval, as well as the performer's interpretation of pitch structure. Greater concentration is required when performing a pitch that is tonally dissonant such as the flat fifth than when performing a consonant interval such as the perfect fifth. For musicians trained in Western music, this effort arises because dissonant pitches occur rarely in Western music and are therefore unstable in memory. Facial expressions may also reflect the performer's perception and appreciation of consonance and dissonance in music, illustrating the emotional impact that such consonance and dissonance has on the performer. Singers also require greater concentration when singing a large rather than a small melodic interval because performing such a large interval involves a large change in muscular position and is associated with a greater potential for pitch errors.

Performers may even dramatize the physical effort involved in order to manipulate perceptions of the music by the audience. Emphasizing the dissonant quality of a pitch may serve as an illustrator, encouraging listeners to appreciate the aesthetic use of dissonance in music, as well as a regulator, confirming for listeners that the use of such dissonance was

intended. Exaggerating the apparent effort involved in producing a melodic interval may expand the perceived size of that melodic interval, whereas underplaying the perceived effort may contract the perceived size of that melodic interval. This effect may be observed with musically trained listeners who can assign a categorical label to intervals. For example, judgments of the size of a melodic interval are significantly influenced by differences in timbre (or 'tonal texture') between the two tones, even for musically trained listeners who possess categorical labels for intervals (Russo and Thompson, in press). That is, even though participants had explicit, formal knowledge of pitch-interval category concepts and labels, their subjective experience of these intervals was not determined by that knowledge.

In Experiment 1, we examined the role of facial expressions in shaping perceptions of musical dissonance. Twenty excerpts were selected from audio-visually recorded performances of B. B. King playing the blues. In ten of the selections B. B. King conveyed a strong sense of dissonance in his facial expression (as exemplified in Figure 4). In the remaining ten selections, his facial expression was more neutral. The level of dissonance conveyed by the sonic element of the music did not constrain our

Figure 4. *An example of B. B. King conveying dissonance though the visual mode*

selection of excerpts. Two groups of participants were presented with the 20 excerpts. One group received audio-only presentation and the other group received audio-visual presentation. We expected that the dissonance conveyed by the visual affect would only influence ratings of dissonance in the audio-visual presentation mode.

Twenty-six participants with limited formal training in music were asked to judge the level of dissonance in musical excerpts. All clips were selected to be within 3 to 6 seconds in length. The performer's facial expression could be described as conveying dissonance for 10 of the excerpts and neutral for the remaining 10 excerpts. A sense of visual dissonance was generally conveyed by wincing of the eyes, shaking of the upper body, and a rolling of the head in a back-swung position. Dissonance was described to participants as occurring when the music sounded discordant (i.e., conflicted or negative) and in need of some sort of resolution. Ratings were made on a 7-point scale, where 1 represented 'low dissonance' and 7 represented 'high dissonance.'

Figure 5 plots dissonance ratings for excerpts accompanied by either facial expressions of dissonance or neutral expressions. Statistical analyses

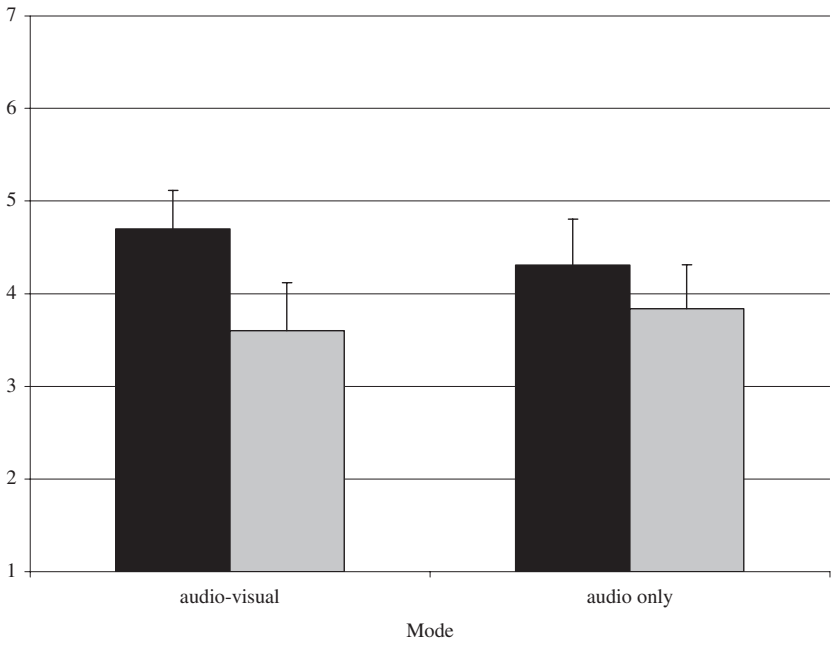

Figure 5. *Mean dissonance ratings (and standard error bars) for intervals presented in audio and audio-visual presentation modes*

indicated a reliable interaction: in the audio-visual presentation mode, but not in the audio-only presentation mode, ratings were considerably higher for performances that were visually dissonant than for performances that were visually neutral (Thompson and Russo [forthcoming]).

In Experiment 2, we examined whether or not a performer was capable of conveying the size of a melodic interval through facial expressions alone. In this experiment, we utilized recordings of intervals performed by a trained singer. Twenty-four participants with limited music training were recruited: one group (n = 12) received audio-only presentation; the other group (n = 12) received video-only presentation. Participants were asked to make judgments of interval size. The performed intervals ranged in size from two to nine semitones. The performer's facial expression and physical gestures served as illustrators, emphasizing the physical effort and dramatic significance associated with a large melodic interval. We expected this visual information to influence the listeners' perception of the size of the melodic interval. The size rating was made on a 5-point scale in which a rating of 1 indicated a small interval and a rating of 5 indicated a large interval.

Figure 6 plots mean size ratings for audio-only and video-only conditions and indicates that listeners were highly sensitive to interval size

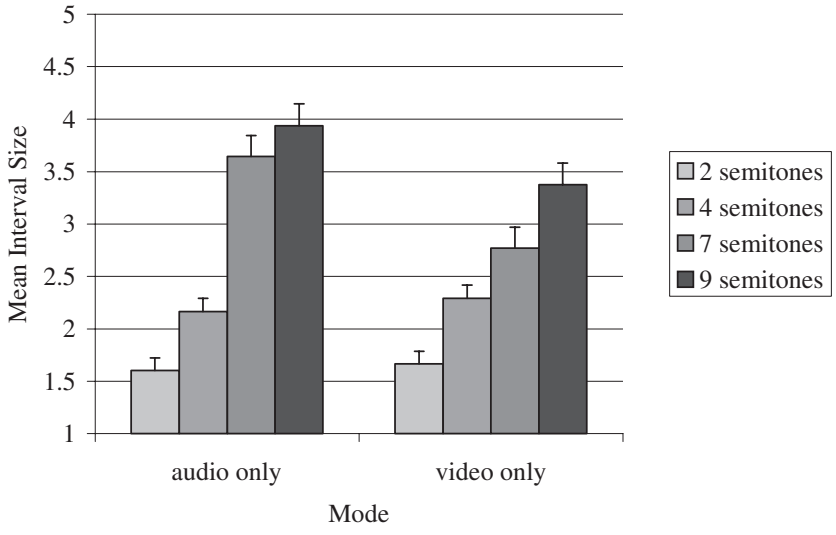

Figure 6. *Mean size ratings (and standard error bars) for intervals presented in audio-only and video-only conditions*

cues conveyed by visual information. In particular, the veridical ordering of interval size was preserved in the absence of any acoustic information.

In Experiment 3, we examined whether visual aspects of music performance influence the perception of interval size when the aural dimension of music is available. Fifteen new participants with limited musical training judged the size of melodic intervals that were either congruent or incongruent with accompanying facial expressions. The audio and video channels of the recordings described in Experiment 2 were recombined to form congruent and incongruent audio-visual presentations. Congruent presentations consisted of audio and visual aspects of the same interval. Incongruent presentations consisted of audio and visual aspects of different intervals. Participants were asked to base their ratings on the audio component alone, and rated interval size using a 5-point scale (1 = small; 5 = large).

As predicted, visual aspects of the sung intervals reliably influenced judgements of interval size. Ratings for audio presentations of a 2-semitone interval were higher when accompanied by the incongruent video (9-semitone interval) than when accompanied by the congruent video. Conversely, ratings for the audio presentation of a 9-semitone interval were lower when accompanied by the incongruent video (2-semitone interval) than when accompanied by the congruent video.

4.2. *Effects of visual information on judgements of affect in music*

All listeners are highly sensitive to emotional meaning in music and performers use a variety of techniques to express such meaning (Balkwill and Thompson 1999; Balkwill, Thompson and Matsunaga 2004; Juslin and Laukka 2003; Sloboda and Juslin 2001; Thompson et al. 2003, 2004). Qualities of music such as tempo, loudness, consonance, mode, and melodic contour provide basic acoustic signals of emotion that can be understood by listeners from an early age (Juslin and Laukka 2003). Emotional meaning can even be understood in unfamiliar music from other cultures (Balkwill and Thompson 1999; Balkwill et al. 2004) and many acoustic cues observed in music are also observed in speech (Balkwill and Thompson n.d.; Ilie and Thompson, in press; Juslin and Laukka 2003; Thompson et al. 2004). In a music performance, emotions are conveyed not only in the sounded events but also in facial expressions and gestures.

In Experiment 4, we presented listeners with video recordings of sung melodic intervals and asked them to judge the affective valence of those intervals. The interval was either a major third, which should convey a

positive emotional message, or a minor third, which should convey a negative emotional message. Two judges independently documented the visual cues (i.e., affect display) available in each performance. Major intervals were performed with raised eyebrows, widening of the eyes, and a slight smile (lips come together); minor intervals were performed with little movement in the eyes, eyebrows, or mouth. In order to isolate the influence of facial expressions, audio-visual presentations were either congruent or incongruent. In a congruent presentation, the intended emotion in both modes was identical (e.g., positive valence). In an incongruent presentation, the intended emotion in both modes was different (e.g., happy audio with sad video).

Twenty-three participants with a wide range of music backgrounds were presented with audio-visual recordings of the sung intervals and were asked to rate the affect being conveyed in each performance on a five-point scale. A rating near 1 indicated that the interval conveyed a negative emotional message, and a rating near 5 indicated that the interval conveyed a position emotional message. As with Experiment 3, participants were told to base their judgments on the audio component alone. Mean affect ratings for all presentations of major and minor third intervals are plotted in Figure 7. For both the major and minor third sung intervals, mean ratings were higher (more positive) if the interval was

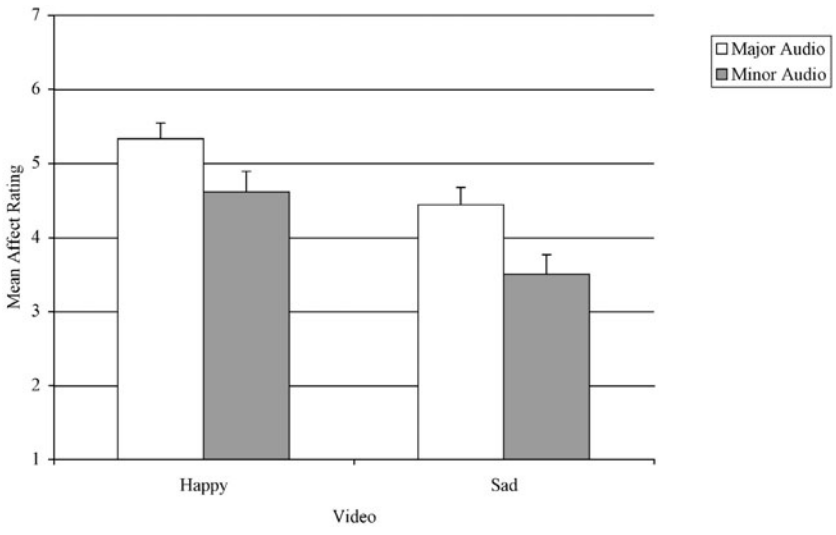

Figure 7. *Mean affect ratings (and standard error bars) for major and minor intervals (audio) with happy and sad facial expressions*

accompanied by facial expressions used to sing a major (happy) interval than if the interval was accompanied by facial expressions used to sing a minor (sad) interval.

In Experiment 5, we investigated the influence of visual aspects of performance on perceived emotional *valence* (i.e., positive or negative) in filmed vocal performances. We selected 30 short clips (roughly 10 second each) from a number of filmed performances by Judy Garland (Jewison 1962, see Figure 8). Clips were selected so as to be representative of her performance style and included emblems, illustrators, regulators, and affect displays. Recognizing that aural and visual aspects of music typically communicate the same affective valence, our aim was to determine how often (if ever) accompanying visual information actually has a significant impact on listeners' affective interpretation of the music. Two groups of participants were recruited: One group of participants received audio-only presentation; the other group received audio-visual presentation.

Thirty-nine participants with limited musical training were asked to rate the emotional valence of the music on a scale from 1 (highly negative) to

Figure 8. *An example of Judy Garland conveying emotional valence through hand gesture and facial expression*

7 (highly positive). Participants were told to consider only the 'music' when assigning their ratings.

Data analysis revealed that in six performances there was a significant difference between valence ratings for the audio and audio-visual presentation modes. (The number of differences expected by chance is between 1 and 2.) The results indicate that the emotional connotation of a music performance that is both seen and heard is not always the same as the emotional connotation of a performance that is merely heard. Valence ratings were higher for audio presentation mode than for audio-visual presentation mode in five of these performances but the trend was reversed in the remaining performance. These findings confirm that visual aspects of performance sometimes alter the perceived valence of the music and that the direction of influence depends on the performance.

The five experiments confirm that visual aspects of performance reliably influence perceptions and interpretations of music. To a surprising extent, facial expression and other bodily movements affect music experience at a perceptual level and an emotional level. Different facial expressions can cause the same musical events to sound more or less dissonant, the same melodic interval to sound larger or smaller, and the same music to sound more or less joyful. In short, listeners integrate visual with aural aspects of performance to form an integrated audio-visual mental representation of music, and this representation is not entirely predictable from the aural input alone.

5. Visual aspects of music in contemporary media

The empirical evidence described above confirms that visual aspects of music significantly influence our experience of music, and yet music is often conceived as a purely aural experience. This view is implicit, for example, in music pedagogy and theory. We now discuss factors that may have contributed to this pervasive assumption and identify recent genres and media forms that are allowing visual components of music to reemerge as an important aspect of music experience.

A significant historic event in the current context is the invention of the radio and gramophone, which isolated the audio mode of music, reinforcing the notion that music experience was a solely aural phenomenon. At the same time, the visual system was isolated with silent movies, which were entirely visual experiences. The first film was itself an experiment to see whether all four hooves of a galloping horse are off the ground at the same time, so the interest was exclusively visual. Whether by accident, technological underdevelopment, or the general historical tendency in

communications since the printing press, at this point in history there is a purely technological splitting of visual and aural technologies. Eventually, piano or orchestra started to accompany silent films, but this accompaniment functioned originally to mask the distracting sound of the projection machine. Musical accompaniment only had to vaguely match the visual content and was often improvised.

For the generations that experienced music primarily through radio or some other electronic listening apparatus, music experience seems, quite naturally, to be an aural experience. Indeed, the vast majority of music listening continues to be an aural experience through radio listening while driving, or through the more recent contraptions such as personal listening devices (e.g., iPods). Nonetheless, when commercial usage of the radio first started to include the broadcasting of music, the reactions of musicians varied from outrage to acclaim. In 1926, Scholes argued that the broadcast medium of radio 'led to the greater *democratization* of music.' He also described a number of 'minor advantages,' including:

> a diminution of *personality* worship. When you no longer see the pianist throwing his arms about, when you no longer see the tenor or baritone rising on his toes to his top note, then let us hope that people will begin to listen to music as music and not as something which is coming to them from a celebrity ... (Scholes 1926: 17)

However, according to the same author, broadcasting also has disadvantages, leading to 'less reading and thinking' and a 'discouragement of home performances' (Scholes 1926: 18–19). Scholes also noted that when you do not see the actual performer, you somehow 'do not enjoy the playing so much' (Scholes 1926: 18). According to Scholes, the most serious problem with radio is that it creates '... a demand for more *popular* kinds of music, and there may come discouragement to serious composers,' with an inevitably 'cheapening influence' (Scholes 1926: 19). Such outrages continued through to the audio-visual era. Bands like *Milli Vanilli* have been embarrassed on world broadcast when they were discovered lip-synching after a playback machine broke down during their performance.

Paradoxically, popular artists today must, above all else, look the part, and the ratio of visual to overall musical content can be seen to have produced an aesthetic economy that threatens to privilege visual over aural aspects of music. This may lead to a stifling of musical forms while promoting a proliferation of visual 'styles' (cf. Fairclough 2000). The emphasis on the visual has led to a number of challenges for musicians: the diminution of music as an *art form* due to the *formulaic* (i.e., generic) demands of visually oriented film and video productions; the emphasis on

(V7(M) 10/5/05 15:56) WDG/G J-1299 Semiotica, 156 PMU: WSL(V) 10/05/05 Times_M (0).3.04.05 (148×225mm) pp. 177–202 007_P (p. 195)

the visual appeal of performers over their talent; and distribution models that exclude many excellent musicians from participating in the so-called *music industry*, to the point at which only 1200 musicians in a country the size of Germany can make a living from music.

The general diffusion of the video clip as a genre for music promotion in the early 1970s entailed an interesting reversal of the usual mediations developed by the powerful film monopolies. Rather than starting with visual information and accompanying it with music, music videos *started* with music production and visual materials were developed to fit the music. In narrative film and video production, the mediation chains work in the opposite direction.

Nonetheless, in spite of visuals being the apparent accompaniment in music videos, the evolution of MTV has shown a gradual reestablishment of visual domination in the economy of musical sensitivities. In early music clips, the performing band was featured heavily, emphasizing the performance and production of the music. Almost immediately, however, producers recognized the need for music stars to be physically interesting, if not conventionally attractive in appearance. The appearance or look of earlier performers such as Hendrix, Joplin, and other such rebel types had a look that made as much a statement of rebellion as their music. As videos evolved, the aesthetic field of visual information expanded, becoming more intricate, complex, expensive, and dominant. In most music videos today, the aesthetic field of visual information is far more expansive and expensive than that of the music. The listener's attention is focused on visual details; their attention to the music is less focused and they are left with only sketchy and generic impressions.

The increased use of visual resources in new media has altered the alignment or co-regulation of perception and affect that occurs between performers and listeners. If part of the performer's interpretation of a piece of music is conveyed through the visual modality and if the visual modality can influence an audience's interpretation of that music, then substituting the gestures of a performer with other visual content necessarily changes perceptual and affective co-regulation, distorting and diluting the communication between performers and listeners through a literal distancing of the performer from her or his audience. Moreover, contemporary strategies for music deployment combine input from multiple participants. The modern music video, for example, not only reflects the interpretation of the performing artist but that of the stylist, director, producer, and other participants in production and post-production. The result is a new form of musical interaction, with participants displaced in space and time. Musical interaction is distributed among performers, producers, and listeners who participate in different times, places and

contexts. The process is filtered and distorted by media that deploy only the aural component, or that overlay new, often irrelevant or distracting visual information onto the aural experience.

The increased use of visual resources in music performance also has political economic implications. Because the visual dimension has been 'reintroduced' to music through new technologies, and because of the pervasive and influential use of visual technologies in music clips and live technologies, it can be argued that the visual aspects of music performance have become more influential than at any other time in history, precisely through their technological separation and reintegration. Massive lighting and public address systems; expensive video production and recording techniques; the proliferation of MTV-type programs to promote musical performers and performances; and the tightly controlled global distribution networks through which these performances are disseminated present enormous barriers to entry for young and emerging musicians in the global music economy.

Visual values have increasingly come to dominate the political economy of sensory appeal. Today, this has resulted in a proliferation of visual 'styles,' or generic 'ways of being' (Fairclough 2000) that are produced for popular music clips. Examples of how this has played out include lip-synched performances, and more recently in the 'air guitar' phenomenon, in which undue attention is drawn to the clichéd, generic gestures and exaggerated movements of lead guitarists. National air guitar competitions are now held in many countries throughout the world, and every year an international competition is held. Air guitar experts pay close attention to every gesture and movement of guitarists, down to individual hand gestures. The judgments are generic; that is, they are responsive to the regularized patterns or genres of what real guitar players do. The air guitar phenomenon suggests that visual-stylistic aspects of performance are valued in their own right, and may therefore influence judgements of what distinguishes good music from bad.[3] The look, the poses that convey attitude, the hand gestures, and the frenetic movements are all emblematic of what these fans have come to value in their musical experiences.

Thus, visual aspects of music remain critical in live performance, but the facial expressions, hand gestures, and other movements of performers are gradually being diluted or even replaced by other kinds of visual information that are presented in popular music experiences. This movement away from the actions of the performer began with the super stadium concerts of the early 1970s, the most obvious exemplar being Pink Floyd, with its high ratio of visual machinery to actual people performing music. Mesmerizing lights shows, visual tricks and surprises, and

(V7(M) 10/5/05 15:56) WDG/G J-1299 Semiotica, 156 PMU: WSL(V) 10/05/05 Times_M (0).3.04.05 (148×225mm) pp. 177–202 007_P (p. 197)

elaborate sets became markers of prestige in themselves. Music videos have continued this trend away from the performance itself, with producers preferring fragmentary sequences of startling images, an emphasis on fantasy and liminal states, and a disruption of traditional narrative. This visual information, as with facial expressions and gestures of performers makes reference to the music and can be interpreted on a semiotic level. But unlike the nonverbal behavior of live musical performers (e.g., emblems, illustrators, affect displays, and regulators), video images can be detached from the aural experience. Rather than consistently supporting the music, these video images may compete for our attention. In this way, performance in music videos — in spite of their innovative use of both audio and visual materials — may reinforce the separation of audio and visual dimensions of music that was begun with the invention of the radio, gramophone, and silent movie.

Notes

1. Halliday (1994) provides a metafunctional distinction between different dimensions of meaning: 'Ideational' meaning ('of'-ness or 'about'-ness of an utterance), 'Interpersonal' meaning (how a meaningful event reflects, changes, or reinforces social relationships), and 'Textual' meaning (how an utterance is textured at multiple levels to provide coherence). These aspects are seen to be part of any and all utterances and happen at the same time. Similarly, Lemke (1995) provides a metafunctional distinction between 'Presentational' meaning (the way elements and aspects of the world are typically represented by a particular community), 'Orientational' or 'Attitudinal' meaning (the attitudes expressed towards these elements and aspects in a given utterance), and 'Organisational' meaning (how coherence is derived through internal and external connections between text and context).
2. An example of dissonance was illustrated by Chuck Jones who, in one Warner Brothers cartoon, had Bugs Bunny play 'If all those Endearing Young Charms' with the final note flattened by one semitone. The correct note had been rigged inside the piano by Yosemite Sam to ignite some TNT upon being played. The repeated dissonance of the flat note eventually aggravates Yosemite Sam to the point at which he is compelled to play the right note, thus blowing himself up.
3. This phenomenon was evident in popular music appreciation when the artist Christopher Cross became an overnight radio hit, unexpectedly selling millions of albums. His company rushed to make a music video, but unfortunately Cross was not photogenic. His record sales stopped and he was never heard of again.

References

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

—(n.d.). Decoding speech prosody in five languages: Canadian and Japanese judges. Submitted to *Emotion*.

Balkwill, L. L., Thompson, W. F., and Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research* 46 (4), 337–349.

Clayton, M. (forthcoming). Communication in Indian raga performance. In *Musical Communication*, D. Miell, R. MacDonald, and D. Hargreaves (eds.). Oxford: Oxford University Press.

Cohen, A. (2001). Music as a source of emotion in film. In *Music and Emotion: Theory and Research*, P. N. Juslin and J. A. Sloboda (eds.), 249–274. New York: Oxford University Press.

Cook, N. (1998). *Analysing Musical Multimedia*. Oxford: Clarendon Press.

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

—(1994). What type of information in conveyed in the body movements of solo musician performers? *Journal of Human Movement Studies* 6, 279–301.

—(1995). What does the visual information contained in music performances offer the observer? Some preliminary thoughts. In *Music and the Mind Machine: Psychophysiology and Psychopathology of the Sense of Music*, R. Steinberg (ed.), 105–114. Heidelberg: Springer.

—(2001). The role of the body in the production and perception of solo vocal performance: A case study of Annie Lennox. *Musicae Scientiae* 5 (2), 235–256.

Davidson, J. and Correia, J. S. (2002). Body movement. In *The Science and Psychology of Music Performance*, R. Parncutt and G. E. McPherson (eds.), 237–253. New York: Oxford University Press.

Di Carlo, N. S. and Guaitella, I. (2004). Facial expressions of emotion in speech and singing. *Semiotica* 149 (1/4), 37–55.

Dissanayake, E. (2001). *Homo Aestheticus: Where Art Comes From and Why*. Seattle: University of Washington Press.

Ekman, P. and Friesen, W. V. (1981). The repertoire of nonverbal behaviour. In *Nonverbal Communication, Interaction, and Gesture*, A. Kenson (ed.), 57–106. Netherlands: Mouton.

Fairclough, N. (2000). Discourse, social theory, and social research: The discourse of welfare reform. *Journal of Sociolinguistics* 4 (2), 163–195.

Gabrielsson, A. (1999). The performance of music. In *Psychology of Music*, D. Deutsch (ed.), 501–602. New York: Academic Press.

Gorbman, C. (1987). *Unheard Melodies: Narrative Film Music*. Bloomington: Indiana University Press.

Graham, P. (2003). Critical discourse analysis and evaluative meaning; Interdisciplinarity as a critical turn. In *Critical Discourse Analysis: Theory and Interdisciplinarity*, G. Weiss and R. Wodak (eds.), 130–159. London: Palgrave MacMillan.

Halliday, M. A. K. (1978). *Language as a Social Semiotic*. Victoria: Edward Arnold.

—(1994). *An Introduction to Functional Grammar*, 2nd ed. London: Arnold.

Husain, G., Thompson, W. F., and Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities: Re-examination of the 'Mozart effect'. *Music Perception* 20, 151–171.

Ilie, G. and Thompson, W. F. (in press). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*.

Jewison, N. (dir.) (1962). *Judy, Frank and Dean: Once in a Lifetime*. Wiley Entertainment.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In *Music and Emotion: Theory and Research*, P. N. Juslin and J. A. Sloboda (eds.), PAGE NUMBERS. New York: Oxford University Press.

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

Kilbourne, J. (1999). *Can't Buy My Love: How Advertising Changes the Way We Think and Feel*. New York: Simon and Schuster.

Kurosawa, K. and Davidson, J. W. (2005). Nonverbal behaviours in popular music performance: A case study of *The Corrs. Musica Scientiae* 9 (1), 111–133.

Lemke, J. L. (1995). *Textual Politics: Discourse and Social Dynamics*. London: Taylor and Francis.

—(1998). Analysing verbal data: Principles, methods, and problems. In *International Handbook of Science Education*, K. Tobin and B. Fraser (eds.), 1175–1189. New York: Kluwer.

Marks, L. E. (1978). *The Unity of the Senses: Interrelations Among the Modalities*. New York: Academic Press.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748.

McPherson, G. and Thompson, W. F. (1998). Assessing music performance: Issues and influences. *Research Studies in Music Education* 10, 12–24.

Pearce, R. (dir.) (2003). The Road to Memphis, written by R. Gordon. From *Martin Scorsese Presents the Blues: A Musical Journey*. Vulcan Productions, Road Movies, in association with Cappa Productions and Jigsaw Productions.

Russo, F. A. and Thompson, W. F. (in press). An interval size illusion: Extra pitch influences on the perceived size of melodic intervals. *Perception and Psychophysics*.

Saldaña, H. M. and Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception and Psychophysics* 54 (3), 406–416.

Scholes, P. A. (1926). Broadcasting and the future of music. *Proceedings of the Musical Association, Fifty-Third Session*, 15–37.

Sloboda, J. A. and Juslin, P. N. (2001). Psychological perspectives on music and emotion. In *Music and Emotion: Theory and Research*, P. N. Juslin and J. A. Sloboda (eds.), 71–104. New York: Oxford University Press.

Thompson, W. F. (2002). Evoking terror in film scores. *M/C: A Journal of Media and Culture* 5 (1). Available online at http://journal.media-culture.org.au/0203/evoking.php.

Thompson, W. F., Russo, F. A., and Sinclair, D. (1994). Effects of underscoring on the perception of closure in film excerpts. *Psychomusicology* 13, 9–27.

Thompson, W. F., Diamond, C. T. P., and Balkwill, L. L. (1998). The adjudication of six performances of a Chopin Etude: A study of expert knowledge. *Psychology of Music* 26, 154–174.

Thompson, W. F., Schellenberg, E. G., and Husain, G. (2003). Perceiving prosody in speech: Effects of music lessons. *Annals of the New York Academy of Sciences* 999, 530–532.

—(2004). Decoding speech prosody: Do music lessons help? *Emotion* 4, 46–64.

Thompson, W. F. and Russo, F. A. (forthcoming). Visual influences on the perception and appreciation of music.

Vines, B. W., Nuzzo, R. L., and Levitin, D. J. (in press). Analyzing temporal dynamics in music: Differential calculus, physics, and functional data analysis techniques. *Music Perception*.

William Forde Thompson (b. 1957) is the Director of the Institute for Communication and Culture at the University of Toronto at Mississauga ⟨b.thompson@utoronto.ca⟩. His

interests include music, gesture, and cognition. His recent publications include 'Decoding speech prosody: Do music lessons help?' (with E. G. Schellenberg and G. Husain, 2004); 'Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners' (with L. L. Balkwill and R. M. Matsunaga, 2004); and 'The attribution of meaning and emotion to song lyrics' (with F. A. Russo, in press).

Phil Graham (b. 1961) is Canada Research Chair in Communication at the University of Waterloo and Senior Lecturer in Communication at the University of Queensland ⟨pwgraham@uwaterloo.ca⟩. His research interest is in the political economy of communication. His recent publications include 'Militarising the Body Politic: New media as weapons of mass instruction' (with A. Luke, 2003); 'Predication, propagation, and mediation: SFL, CDA, and the inculcation of evaluative meaning systems. Systemic functional linguistics and Critical Discourse Analysis' (2004); and 'A call to arms at the End of History: A discourse-historical analysis of George W. Bush's declaration of war on terror' (with T. Keenan and A. Dowd, 2004).

Frank A. Russo (b. 1969) is a Post-Doctoral Research Fellow at the University of Toronto at Mississauga ⟨frusso@utm.utoronto.ca⟩. His interests are in affective and aesthetic aspects of human communication. His recent publications include 'Learning the "Special Note": Evidence for a critical period for absolute pitch acquisition' (with D. L. Windell and Lola L. Cuddy, 2003); 'The subjective size of melodic intervals over a two-octave range' (with W. F. Thompson, in press); and 'An interval-size illusion: The influence of timbre on the perceived size of melodic intervals' (with W. F. Thompson, in press).

(V7(M) 10/5/05 15:56) WDG/G J-1299 Semiotica, 156 PMU: WSL(V) 10/05/05 Times_M (0).3.04.05 (148×225mm) pp. 177–202 007_P (p. 201)