

## Seeing Music? What musicians need to know about vision

MICHAEL SCHUTZ[1]

*University of Virginia  
Department of Psychology*

*Longwood University  
Department of Music*

**ABSTRACT:** Music is inherently an auditory art form, rooted in sound, and generally analyzed in terms of its acoustic properties. However, as the process of hearing is affected by seeing, visual information does in fact play an important role in the musical experience. Vision influences many aspects of music – from evaluations of performance quality and audience interest to the perception of loudness, timbre, and note duration. Moreover, it can be used to achieve musical goals that are in fact acoustically impossible. As such, understanding the benefits of embracing (and the costs of ignoring) vision’s role is essential for all musicians. Furthermore, since music represents a pervasive and ubiquitous human practice, this topic serves as an ideal case study for understanding how auditory and visual information are integrated. Given that some musically-based studies have challenged and even contributed to updating psychological theories of sensory integration, this topic represents a rich area of research, relevant to musicians and psychologists alike.

Submitted 2008 Aug 15; accepted 2008 Aug 29.

**KEYWORDS:** *sensory integration, cross-modal interactions, music perception, music cognition, vision, audition, psychology, performance, education*

“My god, what does sound have to do with music anyway?”  
- Charles Ives , *Essays Before a Sonata*

WHY do we buy concert tickets when the same sounds can be heard more cheaply and comfortably within our own homes? Why do popular music concerts often include elaborate lighting and staging effects for what is ostensibly an auditory event? Why can’t orchestral musicians wear t-shirts and flip-flops? The answer to these and other similar questions is that visual information in fact plays an important role in shaping the musical experience. In addition to increasing overall excitement and interest performers use it strategically, harnessing its communicative power to supplement and augment their acoustic output. Therefore rather than a distraction, it is actually a tool useful for musical communication.

Although the role of visual information in music is receiving increasing attention by researchers (see Thompson, Graham, & Russo, 2005 and Finnäs, 2001), this work is rarely formally disseminated to or discussed by performing musicians. In part, this gulf stems from the fact that performers and researchers often work in different buildings (if not departments or even universities), rarely attend the same conferences, and almost never read one another’s journals. This is not only unfortunate but also ironic, given the obvious importance of this topic to performing musicians (and educators). Therefore the goal of this article is to review research on the role of visual information in music in a manner helpful to performers and researchers alike.

The first section will touch on three topics essential to understanding the role of visual information in music. The second reviews the growing literature on this topic organized by musical level. The third discusses questions raised from this survey with respect to the ability of certain instruments to benefit most from the use of visual information. The fourth and final section sets forth reasons for embracing vision as a useful, irreplaceable, and above all natural part of the musical experience. For the sake of clarity and ease-of-use, this structure is outlined (with section numbers) in the table of contents below.

---

This article is the result of an invitation extended by William Forde Thompson (editor of *Empirical Musicology Review*). Many thanks to him as well as Frank Russo, Tom Banton, Michael Kubovy, Steve Boker, Kristopher Keeton, Brian McNulty, Zachary Cairns, Warren Williams, Bradley Vines, Bill Epstein, Rachel Keen, and Dennis Proffitt for helpful feedback and suggestions on drafts of this article. This work was supported in part by NIDCD (R01 DC 005636, Michael Kubovy, PI)

## Table of Contents

### 1 Introduction and background

- 1.1 Audio-visual Interactions: Understanding the Process
  - 1.1.1 Detecting audio-visual relationships
  - 1.1.2 Distinguishing between perceptual and cognitive influences
- 1.2 Three Approaches to Research: A Matter of Scope
  - 1.2.1 Types of stimuli
  - 1.2.2 Focus of this review
- 1.3 Psychology and Music: A Symbiotic Relationship
  - 1.3.1 Psychology for musicians
  - 1.3.2 Music for psychologists

### 2 Review of the literature

- 2.1 Influences on High-level Musical Properties
  - 2.1.1 Expressive intentions
  - 2.1.2 Audience interest/emotional engagement
  - 2.1.3 Performance quality
  - 2.1.4 Emotional valence
- 2.2 Influences on Intermediate-level Musical Properties
  - 2.2.1 Interval affect
  - 2.2.2 Lyric comprehension
  - 2.2.3 Dissonance
  - 2.2.4 Tension
- 2.3 Influences on Low-level Musical Properties
  - 2.3.1 Timbre
  - 2.3.2 Loudness
  - 2.3.3 Pitch
  - 2.3.4 Note duration
- 2.4 Communication without Influence: Another Approach
  - 2.4.1 Emotional intentions (interpretation)
  - 2.4.2 Expressivity
  - 2.4.3 Pitch distance
  - 2.4.4 Large-scale structure and non-verbal communication

### 3 Instruments and influence

- 3.1 Which Instruments are Best at Using Visual Information?
  - 3.1.1 Voice
  - 3.1.2 Percussion
- 3.2 Other Observations

### 4 Discussion

- 4.1 Relevance for Psychologists and Musicians
- 4.2 The Musical Experience
  - 4.2.1 Acoustics and music: An important distinction
  - 4.2.2 Vision: Augmenting or interfering?
- 4.3 Conclusions

## 1 INTRODUCTION AND BACKGROUND

The aspects of music that can be visually controlled and/or manipulated range from high-level musical attributes such as evaluations of performer expressivity (Davidson, 1993; Davidson, 1994), performance quality (Wapnick, Darrow, Kovacs, & Dalrymple, 1997; Wapnick, Kovacs Mazza & Darrow, 1998), and audience interest (Broughton & Stevens, in press), to low-level musical attributes such as the perception of pitch (Thompson et al., 2005; Gillespie, 1997), loudness (Rosenblum & Fowler, 1991) and timbre (Saldaña & Rosenblum, 1993). Given the degree and extent of vision's role, understanding the particular aspects of music that are affected is essential for performing musicians, audiences, music educators, and researchers interested in studying its practice. It is also informative for psychologists interested in how visual information influences the perception of sound, as it inherently captures important characteristics of the process occasionally lost under the tightly-controlled laboratory conditions favored in psychological research.

The first section of this paper is designed to provide a foundation for the literature review in the second. To this end, it covers three topics:

- 1.1 Why seeing might affect hearing, and the processes by which this occurs
- 1.2 The degree to which traditional psychological research applies to actual musical performances
- 1.3 Rationale for the scope of this review, as exemplified through a recently discovered illusion.

### 1.1 Audio-visual Interactions: Understanding the Process

#### 1.1.1 DETECTING AUDIO-VISUAL RELATIONSHIPS

Single events can often be detected through multiple modalities – for example, a rock dropping on the ground produces both auditory and visual information. Although we experience these sights and sounds as related, this actually represents an impressive achievement on the part of the perceptual system, given that auditory and visual information are received at different times and processed in different areas of the brain. Therefore, one challenge for the perceptual system is to decide which sights and sounds should be bound together so as to give rise to an internal experience best aligned with the external world. This challenge is exacerbated by the fact that sounds are rarely heard in isolation, but rather amidst a noisy background rich in auditory information.

For example, when walking through the woods we may hear birds chirping and twigs breaking, as well as the sounds of our own walking and breathing. Concurrently, we see feathered animals opening and closing their beaks and twigs snapping in two, etc. As auditory and visual information are processed separately, the brain must ultimately make a decision about which sounds should be experienced as originating from which sights. To this end, the brain uses 'identity cues' to determine which auditory and visual information are related and should therefore be experienced as such (Spence, Sanabria, & Soto-Faraco, 2007; Vatakis & Spence, 2007; Vroomen, 1999; Welch, 1999; Welch & Warren, 1980).

An additional benefit of this integration is that sight and sound each offer some information not available to the other. For example, humans are far better at visual (vs. auditory) source-localization. Consequently, the brain has evolved to use visual information regarding localization when processing sound so as to understand events better than if the two were processed independently (as such, audio-visual integration is a perfect example of the whole being greater than the sum of the individual parts). Artists often exploit these principles, "fooling" the perceptual system by taking advantage of this design. For example, movie actors' voices generally appear to originate from their on-screen images, even though they are actually coming from different spatial locations in the theater (an illusion known as the 'ventriloquist effect'). This happens because the brain binds together related auditory and visual information based on the (generally correct) assumption that faces and voices originate from the same point in space. Because this assumption was true throughout the entirety of our evolutionary development, it represents a perfectly reasonable and useful heuristic for maximizing the utility of auditory and visual information.

#### 1.1.2 DISTINGUISHING BETWEEN PERCEPTUAL AND COGNITIVE INFLUENCES

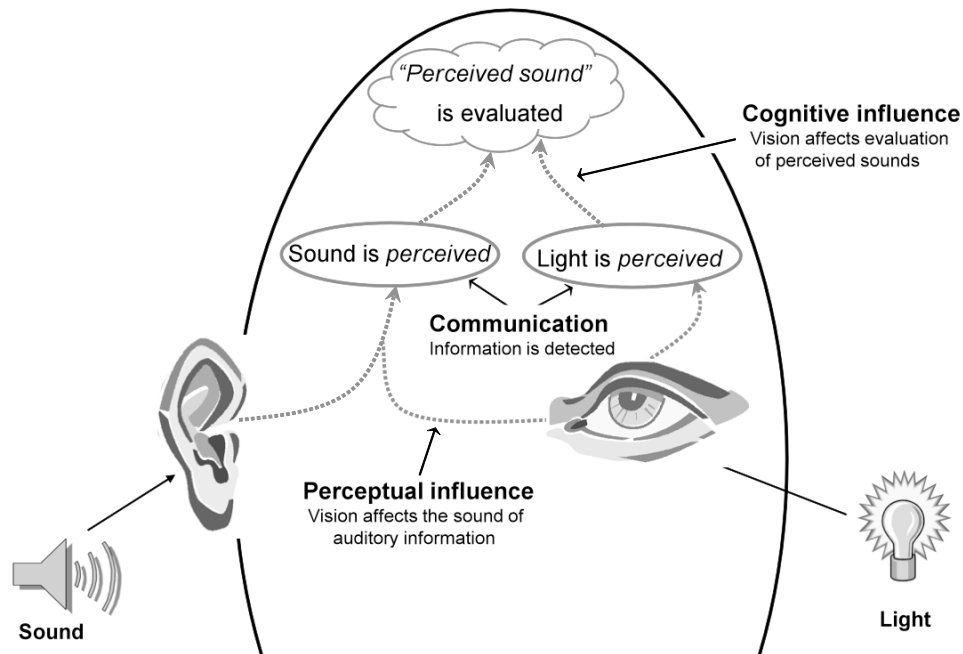
Vision can play an important role in hearing either by (1) influencing the perception of a sound or (2) affecting the conscious evaluation of that sound. Influences that obligatorily alter the internal experience of sound are regarded as perceptual, whereas influences affecting not our experience of sound but rather the

way that sound is evaluated are cognitive. Examples of each are described below, and represented schematically in Figure 1.

*Perceptual influence:* Watching a singer's lip movements improves lyric comprehension by 18% (Hidalgo-Barnes & Massaro, 2007). Musically speaking, such influences are akin to outright acoustic changes – there is no musical difference between articulating song lyrics in a more comprehensible manner and allowing audiences to see lip movements resulting in the same comprehensibility improvement. To meet this classification, integration must be obligatory (e.g. it cannot be ignored), and must not require conscious attention (Arieh & Marks, 2008).

*Cognitive influence:* Evaluations of performance quality are affected by performer attractiveness (Wapnick et al., 1997). Here, visual information changes less the way sounds are perceived than the way those perceptions are evaluated. It is thought that conscious attention is required for cognitive influences, and it may be possible that such influences can be minimized and/or eliminated.

In some cases, a study may show that musical information can be visually detected, but not whether this information actually affects auditory processing. Accordingly, such studies demonstrate not visual *influence* but visual *communication*. For example, the finding that a performer's emotional intentions can be understood by watching their body movements alone (in the absence of auditory information) illustrates the visual communication of musically relevant information (Dahl and Friberg, 2007). However, further work is needed to explore whether such information actually affects the processing of sound.



**Fig 1.** Psychological levels of influence. Visual information can influence either the perception of auditory information (prior to a sound's conscious experience), or the way that sound is cognitively evaluated.

These distinctions are useful for musicians as they comment on *how* visual information shapes the musical experience. For example, although some cognitive influences can be ignored, this is simply not possible for perceptual influences given that they occur prior to our conscious experience of sound. Similarly, it is not yet known whether examples of visual communication actually play any musical role (although future experiments may show this to be the case).

Additionally, this distinction is important for interpreting experiments within the context of previous work. For example, psychophysical research on audio-visual integration (Calvert, Brammer, & Iversen 1998; King & Calvert, 2001) can be useful in interpreting studies exploring perceptual influences such as judgments of pitch, timbre, and loudness. Likewise, the rich social psychology literature on the

role of appearance and attractiveness (Byrne, London, & Reeves, 1968; Clifford et al., 1973; Dion, Berscheid, & Walster, 1972; Lapitsky & Smith, 1981) can be helpful in understanding the effects of clothing and attractiveness on the evaluation of performance quality and emotional engagement (a cognitive influence).

## 1.2 Three Approaches to Research: A Matter of Scope

One important aspect of any literature review is deciding what should not be covered. Given that the goal of this article is to discuss the role of visual information in *music*, it is important to recognize that not all studies involving musical sounds and images are necessarily of practical musical value. For example, pairing a cellist's bowing and plucking motions with both bowed and plucked sounds has shown a visual influence on the perception of timbre (Saldaña and Rosenblum, 1993). Although this demonstrates such influences are possible, it is of limited applied value given that cellists simply cannot use bowing motions to produce plucked timbres (although practical techniques for using this influence may be discovered in the future). To this end, it is useful to recognize the degree to which stimuli used in a given experiment are actually musically realistic.

### 1.2.1 TYPES OF STIMULI

With respect to practical applications, research approaches can be classified into three categories. Differing primarily in the types of stimuli used, they reflect different compromises between musical relevance and experimental precision:

- 1) Stimuli restricted to actual musical performances (or excerpts thereof)
- 2) Stimuli restricted to musical components
- 3) Computer generated stimuli

Each of these research approaches brings a different mix of benefits and drawbacks. The major benefit of the first is that its conclusions are of immediate practical musical value. However its broad scope makes it difficult to draw the types of specific conclusions useful in fully applying such results to actual musical practice. For example, although Davidson (1993) has shown that body movements play a large role in helping audiences understand a performer's expressive intentions, it is not clear which *particular* movements are responsible for these effects, nor the ways in which they convey specific expressive intentions.

To this end, it is sometimes advantageous to use unrealistic combinations of stimuli to answer practical musical questions. For example, to explore the role of body movements in further detail, Thompson, Russo and Quinto (in press) examined the effects of facial information on auditory perception by recording a singer performing intervals of various sizes, then swapping the auditory and visual components. Here, sidestepping the limitations of reality led to clear evidence that facial information can influence the perception of musical intervals. Yet such approaches also have limitations – given that they are based on real-world sounds they cannot be manipulated beyond the ability of a given performer. Furthermore the difficulty of controlling for variability in performer ability and style, microphone quality and sensitivity, and the recording environment used means they can be both time-consuming to create and tedious to equate.

Consequently, artificially generated stimuli hold some important benefits. They are simple, convenient, and not limited to the constraints of our world. Their duration, frequency, and intensity can be easily and precisely controlled. Accordingly, artificial sounds are now popular in many psychologically-based investigations of auditory-visual integration (Alais & Burr, 2004; Walker & Scott, 1981; Welch & Warren, 1980). However, the experimental precision gained from this approach comes at a cost. Given that such sounds are far removed from those used in musical contexts, experiments based on them do not necessarily reflect the ways in which musical sounds (or in some cases natural sounds in general) are processed.

### 1.2.2 FOCUS OF THIS REVIEW

Because of the musical goals of this paper, it is focused primarily on stimuli exemplifying the first approach – those representative of realistic musical conditions. However, experiments using the second (unrealistic combinations of realistic musical sounds and images) will also be included for three reasons.

First, chamber ensembles may coordinate among multiple performers to divert attention such that audiences simultaneously experience sounds along with gestures/movements not used in their production [2]. Second, they suggest theoretically possible influences that may in the future be practically applied. And third, the popularity of visual supplements used in some live (video close-ups and/or light shows) and recorded (e.g. music videos) performances means that under limited circumstances, it is possible for performers to move beyond the practical constraints of realistic pairings.

Studies based upon the third experimental approach involving computer generated stimuli will not receive the same attention as their results are less clearly related to the processing of musical sounds. This issue is illustrated by some of my work exploring the role of visual information in the perception of note duration. It also serves as an example of one way in which music and psychology mutually enhance our understanding of appreciation for one another.

### 1.3 Psychology and Music: A Symbiotic Relationship

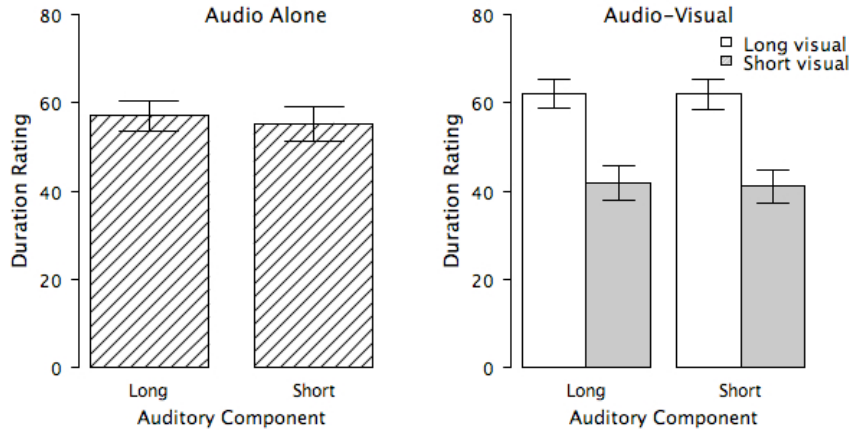
Given that music has evolved in response to the structure of the mind (Huron, 2001), psychology by definition plays an important role in understanding its practice. For example, it provides a helpful framework for resolving a long standing debate among percussionists (Bailey, 1963; Saoud, 2003) – does gesture length (i.e. the up-down-up movement used to strike an instrument) affect note duration? In other words, do physically longer striking gestures produce longer musical notes? Given the compelling nature of the McGurk effect – in which seeing lip movements categorically alters perception of spoken sound (McGurk & MacDonald, 1976) – it is possible that that this disagreement might in fact reflect similar principles.

#### 1.3.1 PSYCHOLOGY FOR MUSICIANS

To tease apart the relative contributions of auditory and visual information to the perception of note duration, Schutz and Lipscomb (2007) video recorded a percussionist playing single notes on the marimba (a tuned, wooden percussion instrument similar to the xylophone) using long and short gestures. To ensure musical relevance, world-renowned performer Michael Burritt (currently the Director of Percussion Studies at the Eastman School of Music) played in a recital hall under conditions similar to those of an actual musical recital. The auditory and visual components were then swapped to create two incongruent conditions (long-gesture/short-audio, and short-gesture/long-audio) supplementing the original congruent conditions (long-gesture/long-audio and short-gesture/short-audio). Participants – music majors at Northwestern University – were asked to judge note duration under two manners of presentation: audio-visual and audio-alone. In the audio-visual condition, participants were asked to base their duration judgments on the sound alone, ignoring the visual gesture.

Two results illustrate the basis for confusion over the relationship between gesture length and musical note duration. First, ratings made in the audio-alone condition did not differentiate between notes produced using long and short gestures (a separate analysis found their durations to be acoustically indistinguishable). Second, ratings in the audio-visual condition were significantly longer when a given note was paired with a long, rather than a short gesture (Figure 2). Therefore, although gestures cannot change the sound of a marimba note, they are capable of changing the way that note *sounds* within the mind of the listener.

This natural audio-visual illusion is particularly useful for percussionists who (unlike other instrumentalists), have minimal control over acoustic note duration independent of other parameters such as volume. Therefore this visual “trick” (which is likely as old as percussion instruments themselves), offers a means for overcoming a major acoustic limitation. Curiously, these results run contrary to the generally accepted notion that visual information does not alter auditory judgments of tone duration – in fact, an informed reading of research on this topic suggests that such results are not possible!



**Fig. 2.** Gestures change perception of marimba note duration. Notes produced using long and short gestures were indistinguishable when presented as audio-alone (left panel). However, the same notes were rated significantly longer when paired with a long gesture than when paired with a short gesture (right panel). Error bars represent 95% confidence intervals. Regenerated from data reported by Schutz and Lipscomb (2007).

### 1.3.2 MUSIC FOR PSYCHOLOGISTS

The artificial tone beeps and light flashes favored by psychologists exploring low-level sensory integration suggest that visual information cannot alter the auditory perception of note length (Welch & Warren, 1980). For example Walker & Scott (1981) report that flashes of light do not affect duration judgments of simultaneously presented tones. Further research has shown that the crucial difference between the Schutz and Lipscomb stimuli and those traditionally used in sensory integration research is the clear causal relationship between modalities – one arising naturally from real-world stimuli. Manipulations breaking the relationship by substituting the sound of the marimba with that of a sustained instrument such as the clarinet or human voice (which could not have been produced by the depicted gesture) fail to show the illusion (Schutz & Kubovy, in press). In other words, the brain appears to be using different integration rules depending upon whether audio-visual pairings are natural (i.e. were encountered over our evolutionary history) or artificial (i.e. encountered mainly in laboratories). Psychologists traditionally focus overwhelmingly on artificially generated sounds (the third research approach), which could not have been caused by realistic gestures (Schutz & Kubovy, 2008). As a result, much psychological research has overlooked a crucial principle governing sensory integration heuristics – cross-modal causality.

This is one illustration of the benefits of using musical stimuli to enhance our understanding of psychology. As music is a human universal practiced in all known cultures and at all known times (Levitin, 2006), conclusions drawn from experiments based on it are more likely to explain how the perceptual system works in the real world than experiments using artificial beeps and flashes. Much as we have evolved as a reflection of our environment, music has evolved as a reflection of the psychological constraints of the human mind (Huron, 2001). Therefore experiments built around it by definition capture important characteristics of the mind's real-world behavior, profitably trading the fine-grained control of artificial stimuli for sounds more representative of those the system actually encounters outside the laboratory. Because of such clear differences between the processing of artificially-generated and musical sounds, this review will focus heavily on the latter. For reviews of research involving artificial sounds, see Calvert et al. (1998) or King and Calvert (2001).

## 2 REVIEW OF THE LITERATURE

Table 1 lists a variety of musical properties and types of evaluations that can be affected by visual information, grouped by musical level. The table is organized into four sections:

- Section 1: Influences on high-level musical properties* – Expressive intentions, audience interest, emotional engagement, performance quality, and emotional valence.
- Section 2: Influences on intermediate-level musical properties* – Pitch-interval affect, comprehension of sung lyrics, musical dissonance, and judgments of tension.
- Section 3: Influences on low-level musical properties* – Timbre, loudness, pitch, and note duration.
- Section 4: Visual communication* – Studies showing that vision *can* convey information rather than examining *how* this information affects the processing of sound.

The six columns of the table indicate:

- Column 1: Aspect influenced* – the type of musical property or dimension influenced. Note that in the last section (listing studies involving only visual information), this column instead indicates the type musical information conveyed.
- Column 2: Musical level* – the musical level of vision’s role, a schema useful in organizing and understanding this literature. Endnote [3] contains the criteria for classifying high, intermediate, and low musical levels.
- Column 3: Psychological level* – the level at which this influence occurs [4]. The listing distinguishes between perceptual and cognitive influences, or is marked “unknown” for experiments in which this is not determinable. Note that for the fourth section (visual communication), there is by definition no influence to classify.
- Column 4: Performer control* – the type of control available over these manipulations within the context of actual performances. For studies using realistic stimuli, control can be regarded as either “direct” (e.g. manipulating body language to convey expressive intention) or “indirect” (e.g. attractiveness affects ratings of performance quality). For studies using only visual information or involving only artificial cross-modal pairings, performer control in hypothetical musical contexts is classified as either “probable” (likely to apply) or “unknown” (unclear whether it could be practically employed).
- Column 5: Instrument used* – For papers or experiments involving different instruments [5], each is listed separately (such as Dahl and Friberg’s 2007 experiments using similar approaches to examine the body movements of a marimbist, bassoonist, and saxophonist).
- Column 6: Experiment* – Citation for the example. For cases in which a particular entry applies to a single experiment within a larger series of studies, the experiment number is included as well. Note that due to space considerations, papers with more than two authors are cited using only the name of the first author.



**Table 1.** Visual influence as a function of musical property/dimension. See previous page for explanation of its structure.

Aspect influenced	Musical level [3]	Psychological level	Performer control	Instrument used	Experiment
Expressive intentions	High	Cognitive	Direct	Violin	Davidson (1993), Exp. 1
	High	Cognitive	Direct	Piano	Davidson (1993), Exp. 2
	High	Cognitive	Direct	Marimba	Broughton and Stevens (in press)
Audience interest	High	Cognitive	Direct	Marimba	Broughton and Stevens (in press)
Emotional engagement	High	Cognitive	Indirect	Piano	Timmers et al. (2006)
Performance quality	High	Cognitive	Indirect	Marimba	McClaren (1988)
	High	Cognitive	Indirect	Voice	Wapnick et al. (1997)
	High	Cognitive	Direct	Violin	Wapnick et al. (1998)
	High	Cognitive	Direct	Piano	Wapnick et al. (2000)
	High	Cognitive	Indirect	Piano	Wapnick et al. (2004)
Emotional valence	High	Cognitive	Direct	Voice	Thompson et al. (2005), Exp. 5
Interval affect	Intermediate	Perceptual	Unknown	Voice	Thompson et al. (in press)
Lyric comprehension	Intermediate	Perceptual	Direct	Voice	Hidalgo-Barnes and Massaro (2007)
Dissonance	Intermediate	Unknown	Direct	Guitar	Thompson et al. (2005), Exp. 1
Tension	Intermediate	Unknown	Direct	Clarinet	Vines et al. (2006)
Timbre	Low	Perceptual	Unknown	Cello	Saldaña and Rosenblum (1993), Exp. 1
Loudness	Low	Perceptual	Unknown	Clapping	Rosenblum and Fowler (1992), Exp. 2
Pitch accuracy	Low	Cognitive	Indirect	Violin	Gillespie (1997)
	Low	Cognitive	Indirect	Viola	Gillespie (1997)
Pitch distance	Low	Unknown	Unknown	Voice	Thompson et al. (2005), Exp. 3
Note duration	Low	Perceptual	Direct	Marimba	Schutz and Lipscomb (2007)
<b>Aspect conveyed</b>					
Emotional intentions	High	n/a	Probable	Marimba	Dahl and Friberg (2007), Exp. 1
	High	n/a	Probable	Bassoon	Dahl and Friberg (2007), Exp. 2
	High	n/a	Probable	Saxophone	Dahl and Friberg (2007), Exp. 2
Expressivity	High	n/a	Direct	Piano	Davidson (1994), Exp. 2
Pitch distance	Low	n/a	Unknown	Voice	Thompson and Russo (2007)
Structure / emotion	High	n/a	Direct	Orchestra	Krumnhansl and Schenck (1997)
Stage presence, inter-performer coordination, extra-musical concerns	High	n/a	Direct	Voice	Davidson (2001)
	High	n/a	Direct	Voice	Kurosawa and Davidson (2005)

This table lists only experiments published in peer-reviewed journals, and does not contain conference proceedings and/or abstracts. Studies elaborating on existing work by providing converging evidence for previously known musical results are not listed in the table separately, but are mentioned during the relevant summaries in sections 2.1.1-2.4.3. Consequently, fewer studies appear in the table than are mentioned throughout this article. To view samples of actual stimuli used in these experiments, see Appendix II for information regarding online supplements.

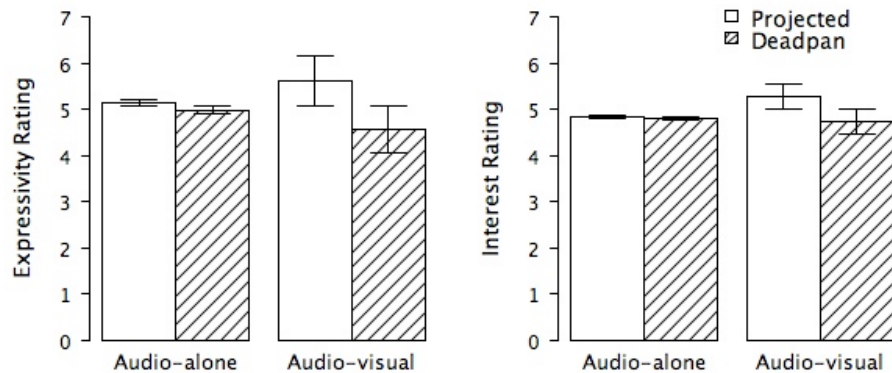
## 2.1 Influences on High-level Musical Properties

### 2.1.1 EXPRESSIVE INTENTIONS

The expressive intentions of a performer may in fact be most clearly detected by watching, rather than by listening. To demonstrate this, Davidson (1993) recorded four conservatory violin students performing musical excerpts of their choice in one of three manners: “deadpan,” “projected,” and “exaggerated.” Undergraduate music students then judged each performance’s level of expressivity on a 7 point scale (with endpoints labeled “deadpan” and “exaggerated”) under three presentation conditions: audio-alone, video-alone, and audio-visual. Participants were best at detecting the intended performance manner in the video-alone condition and worst in the audio-alone condition (although they were still able to differentiate). A second experiment using similar techniques involving a pianist playing an excerpt of the promenade theme from Mussorgsky’s *Pictures at an Exhibition* led to similar results (subsequent investigation of these same videos in experiment 2 of her 1994 paper suggest differences in expressivity ratings here likely reflect differences in the amount of overall body movement used in each condition). Given that discriminability in the audio-visual condition fell somewhere in between that of the audio-alone and video-alone, Davidson went so far as to suggest that vision represents a more efficient mechanism for communicating expressivity, particularly for non-musicians (1995).

Davidson’s results are even more surprising given that participants did not see a full video of the performances, but rather a point light reduction showing only small bands of light tracking the performer’s head, shoulders, hips, elbows, wrists, knees, and ankles. Although humans are adept at reconstructing movement patterns from such displays (Johansson, 1973), it is important to note that participants were not able to observe the performer’s facial expressions or any nuanced movements beyond those of the specified points. The fact that the visual detection of performer intention was even possible (let alone superior!) despite this asymmetric degradation suggests that vision plays an important role in actual musical performances, where audiences are able to observe in much finer visual detail (for similar findings based on videos rather than point-light displays, see Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2005).

### 2.1.2 AUDIENCE INTEREST/EMOTIONAL ENGAGEMENT



**Fig. 3.** Ratings of performer expressivity and participant interest. Expressiveness ratings distinguished between performance manner (left) only in the audio-visual condition. Participants considered the “projected” manner to be more interesting than the “deadpan” manner only in the audio-visual condition (right). Error bars represent  $\pm 1$  standard error. Regenerated from data reported by (Broughton and Stevens, in press).

Extensions to Davidson’s work suggest that such influences occur even when listening to professional musicians, and that they extend to other factors beyond expressive intention. Broughton and Stevens (in press) recorded two professional marimbists playing eight excerpts from the standard contemporary literature in two of the performance manners used by Davidson – “projected” (as if performing in a public recital) and “deadpan.” Although full body motion was available, facial expressions were masked from viewers. Participants were asked to judge performer expressiveness as well as their level of interest under two conditions (audio-visual and audio alone). Ratings of performer expressivity were consistent with

those reported by Davidson – intended performance manner was more easily distinguishable in the audio-visual relative to the audio-alone condition (Figure 3, left panel).

Additionally, they found that the effect of performance manner on audience interest was more salient in the audio-visual condition (Figure 3, right panel). Given that performers rarely attempt deadpan interpretations in actual performances, these results are consistent with McClaren's work suggesting that audio-visual solo performances are judged more favorably (1988). It also agrees with findings suggesting that visual information helps to emotionally engage audiences (Timmers, Marlot, Camurri, & Volpe, 2006).

### 2.1.3 PERFORMANCE QUALITY

In addition to gestures and body language (over which performers retain some moment-to-moment control), incidental information such as physical attractiveness, clothing choice, and stage presence can significantly affect evaluations of performance quality. Although this type of visual information is not context sensitive (i.e. performers cannot change clothing to match their expressive intentions for each section of a concerto), for better or for worse they nonetheless play an important role in shaping the musical experience of audiences and adjudicators. Wapnick et al. (1997) video recorded vocalists performing a song of their choice, then showed these recordings to music faculty and students under one of three conditions: audio-alone, video-alone, and audio-visual. Participants in the video-alone condition evaluated performer attractiveness, whereas participants in the other two conditions evaluated performance quality (specifically diction, intonation accuracy, and musical phrasing, Ekholm, 1994). The male performers [6] rated as more attractive by the video-alone group received higher ratings regarding performance quality from participants in the audio-visual, but not participants in the audio-alone condition. This influence was not offset by training, as faculty and student ratings did not differ.

A similar approach based on performances from the 11th Cliburn International Piano Competition suggests that such influences are found even when evaluating highly accomplished performers. Music students and music faculty (both pianists and non-pianists) from three universities judged one-minute excerpts taken from the competition. Evaluators were divided into two groups, making their ratings under either audio-visual or audio-alone conditions. Ratings given by non-pianists were higher in the audio-visual condition than the audio-alone condition, however this difference was not observed in the ratings given by pianists (Wapnick, Ryan, Lacaille, & Darrow, 2004). Other studies have shown that stage behavior and clothing choices can also influence ratings of performance quality (Griffiths & Davidson, 2006; Wapnick et al., 1998; Wapnick, Kovacs, Mazza, & Darrow, 2000). Together, such studies suggest vision's influence on evaluations of performance quality to be widespread, deeply rooted, and oftentimes unaffected by general musical exposure (although it may be mediated by instrument-specific training).

### 2.1.4 EMOTIONAL VALENCE

A piece's emotional valence (i.e. its "positiveness" or "negativeness") can be a salient aspect of musical performances. To examine whether visual information plays a role in its appraisal, Thompson et al. (2005, experiment 5) asked participants to rate the emotional valence of excerpts from Judy Garland videos using a 7 point scale (ranging from 'highly negative' to 'highly positive'), under either audio-visual or audio-alone manners of presentation. Emotional valence ratings for six of the thirty clips differed based upon presentation condition (significantly more than would be expected by chance). Although it is difficult to quantify how and to what degree vision affected these ratings, from this study it is clear that it can contribute to judgments of emotional valence.

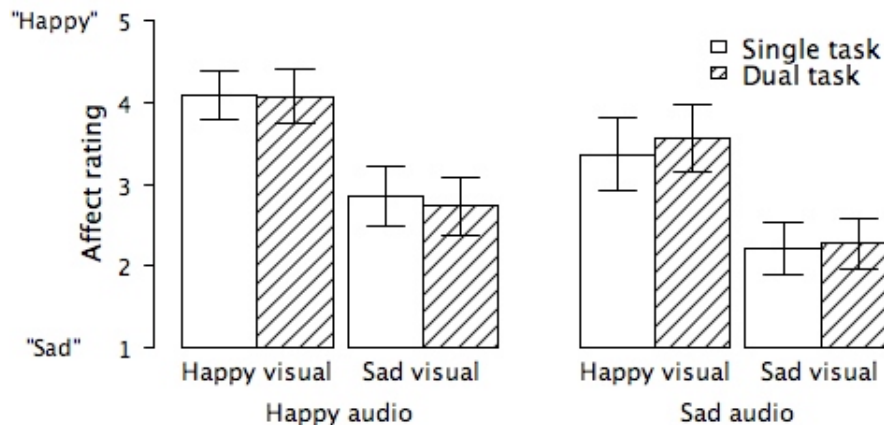
## 2.2 Influences on Intermediate-level Musical Properties

Several types of influences defy classification into strict categories of "low-level" and "high-level" musical characteristics. Although this distinction is useful for organizing the literature, it is more important to recognize vision's pervasive role across these levels than to split hairs over the precise categorization of specific demonstrations. Therefore, studies not clearly falling into either of the previously discussed categories will be regarded simply as showing an influence on "intermediate-level" musical properties.

### 2.2.1 INTERVAL AFFECT

Facial expressions can significantly affect our perception of an interval's affect (essentially its "emotional tone"). In western music, major thirds are frequently described as sounding "happy" and minor thirds as sounding "sad" (Gagnon & Peretz, 2003; Huron, 2006; Kastner & Crowder, 1990). By crossing the auditory and visual information from video-recordings of singers performing major and minor thirds, Thompson et al. (2005, experiment 4) were able to show that visual information can alter the perception of an interval's affective quality (e.g. happy vs. sad). As with many experiments of this type, these results can be interpreted in two ways: (a) visual information altered auditory perception of interval affect; (b) visual information did not alter perception (e.g. the intervals sounded the same), but instead influenced only responses. Given the difficulty of distinguishing between perceptual and decisional shifts (Arieh & Marks, 2008), this issue is both important and difficult to address.

To clarify, Thompson et al. (in press) added two key elements to the original design. First, they instructed participants to base affect responses on the audio alone, discouraging any attempts to consciously integrate across modalities. Second, they added a visual distracter task – monitoring a series of translucent ones and zeros superimposed over the singer's face. Participants were asked to report the number of zeros in addition to judging interval affect. As this secondary task required significant attentional resources (31% of participants failed to perform adequately), it should diminish any visual influence requiring conscious attention. Not only did they replicate their original findings despite these changes, the effect was undiminished by the distracter task (Figure 5). Therefore, this cross-modal integration of affective information is unlikely to depend on conscious attention.



**Fig. 5.** Affect judgments of major and minor thirds. Ratings (5= "happy"; 1= "sad") were influenced by facial movements despite instructions to base responses on the auditory information alone. The degree of influence did not differ between the single task (rate affect) and dual task (rate affect in addition to counting the "zeros") conditions. Error bars indicate 95% confidence intervals about the mean. Regenerated from data reported by Thompson et al. (in press).

### 2.2.2 LYRIC COMPREHENSION

Seeing a vocalist's lip movements can also be helpful in comprehending song lyrics, much as they boost the intelligibility and comprehension of speech (Abry, Cathiard, Robert-Ribes, & Schwartz, 1994; Massaro & Stork, 1998; Middleweerd & Plomp, 1987; Sumbly & Pollack, 1954). Hidalgo-Barnes and Massaro (2007) presented lyrics from thirty-four verses of Primus's *The Pressman* (1993) under three conditions – audio-alone, audio-visual, and visual-alone. For the latter two, visual information was supplied by Baldi, an animated talking head. Lyric comprehension (as identified by participant transcription) was 18% higher in the audio-visual (relative to the audio-alone) condition. Although these effects were smaller than those generally observed in speech, they nonetheless demonstrate the importance of visual information in enhancing audience comprehension of song lyrics.

### 2.2.3 DISSONANCE

Facial expressions can be used to manipulate sounds produced by non-vocal sources such as the guitar. Thompson et al. (2005, experiment 1) asked participants to rate the level of dissonance of 20 musical clips taken from performances given by B.B. King, presented under one of two conditions – audio-visual or audio-alone. The audio-visual group rated clips in which King’s facial expressions visually conveyed dissonance to be more dissonant-sounding than clips in which his facial expressions were neutral. However, the audio-alone group (listening to the same clips) did not judge the level of dissonance to differ significantly, indicating that King’s facial expressions supplemented his acoustic output, presumably clarifying his musical intentions.

### 2.2.3 TENSION

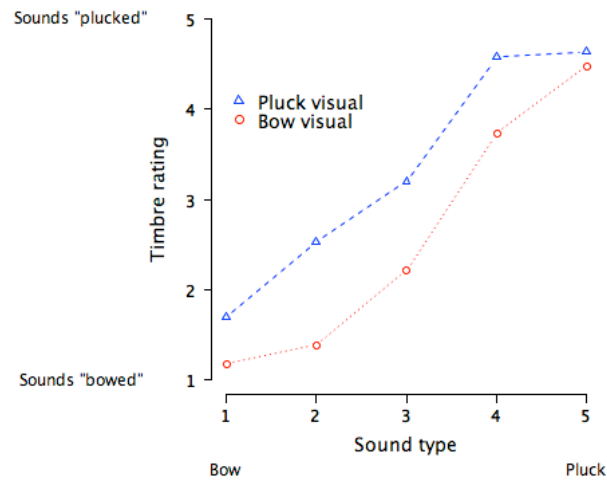
Finally, ancillary gestures – those that are not required for sound production (Wanderley, Vines, Middleton, McKay & Hatch, 2005; Wanderley & Vines, 2006), can be used to manipulate musical tension as well as underscore structural properties. This was shown by asking participants to rate the tension and phrasing displayed in recordings of Stravinsky’s second piece for solo clarinet under one of three conditions: audio-alone, audio-visual, and video-alone. Tension ratings differed (at times significantly) between the audio-visual and audio-alone conditions, indicating that body movements can affect its evaluation (Vines, Krumhansl, Wanderley, & Levitin, 2006). Additionally, judgments of phrasing made in the audio-alone and video-alone conditions exhibited a great deal of similarity, indicating that structural information can be communicated effectively through both modalities (similar to the results of Krumhansl & Schenck, 1997, discussed below). Subsequent experiments using psychophysiological measures such as electrodermal activity to objectively measure responses to these same stimuli are consistent with these conclusions (Chapados & Levitin, in press).

## 2.3 Influences on Low-level Musical Properties

High-level constructs such as “performance quality” and “emotional engagement” are by definition complex, involving a variety of factors. Therefore the fact that visual information serves as one of these is not wholly surprising (although the *extent* of its influence may be). However, given that low-level properties such as timbre, loudness, pitch and note duration are clearly specified by acoustic information, does it not follow that their perception should be unaffected by vision?

### 2.3.1 TIMBRE

By crossing the bowing and plucking movements used by cellists, Saldaña & Rosenblum (1993) showed that visual information plays a role in the perception of timbre. They paired five sounds (ranging from that of a clear bow to that of a clear pluck) with gestures depicting bowing and plucking motions. Participants were informed that the auditory and visual information were not always congruent, and asked to indicate their perception of each sound’s timbre based on the auditory information alone. Their results indicate a significant visual influence (Figure 4) that was greatest for sounds in the middle of the continuum, reflecting that cross-modal interactions are accentuated by ambiguity (as shown by Ernst & Banks, 2002).



**Fig. 4.** Visual influence on the perception of timbre. Timbre judgments (y-axis) of five sounds plotted on the x axis, ranging from a clear bow (left) to a clear pluck (right). The ratings indicate perception of timbre when the sounds were paired with plucking (blue triangle) and bowing (red circle) gestures. Regenerated from data reported by Saldaña and Rosenblum (1993).

### 2.3.2 LOUDNESS

Gestures can also affect the perception of loudness. When crossing the sights and sounds of clapping motions, Rosenblum and Fowler (1992) found that sounds paired with large clapping motions to be judged louder than when paired with the smaller motions used to produce softer claps, despite asking participants to ignore visual information when rating loudness. Although somewhat removed from a musical context, it is likely that such results would replicate with percussive instruments, given that claps are percussive sounds (and have even been used as the basis for entire musical compositions such as Steve Reich's *Clapping Music*). Given that percussionists naturally vary gesture length to control volume (Dahl, 2000; Dahl, 2004), it is likely that such influences already play a role in actual musical performances.

### 2.3.3 PITCH

Judgments of both pitch distance and intonation can be affected by visual information. For example, adjudicator ratings of violin and viola vibrato pitch-accuracy were higher when made under audio-visual rather than audio-alone conditions (Gillespie, 1997). Although the evaluators were not given explicit instructions to ignore visual information, this influence is nonetheless surprising. Given their status as professional musicians teaching at the university level, it seems reasonable to expect they would naturally fixate on auditory information when judging a musical property as important and low-level as pitch accuracy. Although watching the gestures likely distracted attention from intonation (rather than shifting pitch perception directly), it nonetheless demonstrates the ability of visual information to affect judgments of pitch accuracy.

Thompson et al. (2005, experiment 3) report another influence related to pitch – judgments of interval size. When crossing the facial expressions and sounds from small (2 semi-tone) and large (9 semi-tone) sung intervals, visual information significantly influenced judgments such that small auditory intervals were judged to be larger when paired with the larger facial movements, and vice versa.

### 2.3.4 NOTE DURATION

Finally, as discussed previously (section 1.3), percussionists use changes in physical gesture length to control audience perception of musical note duration on the marimba (Schutz & Lipscomb, 2007). This finding is particularly useful, given that it shows a way for percussionists to overcome an acoustic

limitation of their instrument – the inability to manipulate note duration independent of other factors such as volume.

## 2.4 Communication without Influence: Another Approach

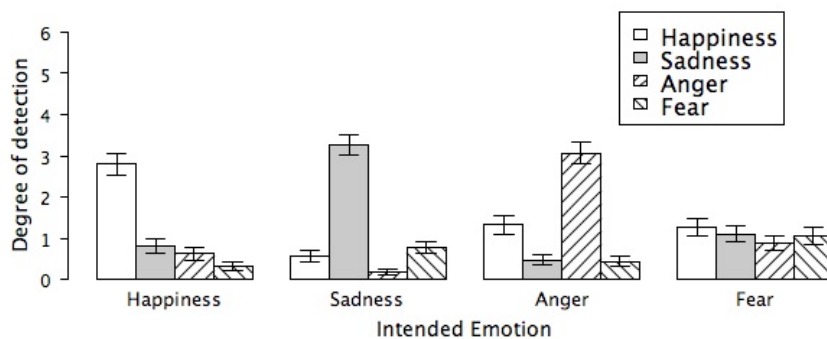
Rather than examining vision's influence on hearing, another approach is to explore the types of musical information that can be communicated through purely visual means. Although it may initially appear counter-intuitive for a musically based study to ignore auditory information altogether, such approaches are desirable for answering certain musical questions. For example, understanding the communicative value of gestures (apart from their acoustic consequences) is most straight-forward when examining ratings of the gestures in isolation. Attempts to analyze ratings of gestures presented with their accompanying auditory information would undoubtedly be confounded by inevitable acoustic differences.

Consequently, studies based on visual information alone can in some cases yield the clearest picture of: (a) the specific types movements employed by performers and dancers (b) the degree to which these movements can be differentiated and recognized by audiences, and (c) the types of information communicated. Armed with this more specific knowledge, it is then possible to further explore how this information can be used to strategically influence the perception and/or evaluation of sound.

### 2.4.1 EMOTIONAL INTENTIONS (INTERPRETATION)

To investigate the role of body movements in communicating emotional intentions, Dahl and Friberg (2007) video-recorded a professional percussionist playing a marimba solo with four different expressive intentions: happiness, sadness, anger, and fear. The videos were filtered (removing facial expressions) and then evaluated by participants according to the degree to which they conveyed four emotions (happiness, sadness, anger, fear) on a scale from 0 (“nothing”) to 6 (“very much”). Their results indicate that participants were quite accurate at recognizing the happy and sad intentions, and were fairly accurate at recognizing the angry intention (although this was often confused with happiness), but were not able to reliably recognize the fearful intention. A second experiment based on the movements of a professional bassoonist and saxophonist reached similar conclusions. Combined results from these experiments (summarized in Figure 6), suggest that expert performers exploit generalized mechanisms for visually communicating their interpretative intentions.

Dahl and Friberg also analyzed how movement patterns vary as a function of expressive intentions, finding that performers used slow and smooth movements to convey sadness, large and fast movements for happiness, and jerky movements for anger. This consistency is interesting, given that performers were asked only to convey a specific expressive intention, and were not informed that analysis would be based on gestures, rather than sound. Therefore, in all likelihood these effects are even stronger in actual musical performances, as they offer significantly more visual information than the filtered videos used here.



**Fig. 6.** Visual detection of performance intention. Participants accurately recognized happiness, sadness, and anger, but were not able to visually detect fearful intentions. Ratings are collapsed across those of the marimbist, saxophonist, and bassoonist. Error bars represent 95% confidence intervals about the mean. Regenerated from data reported by Dahl and Friberg (2007).

#### 2.4.2 EXPRESSIVITY

Continuing her previous work on the visual communication of musical information, Davidson (1994) used point-light recordings to explore the specific areas of a pianist's body that convey expressive information. To do so, she asked a professional pianist to play Beethoven's *Bagatelle No. 11 in Bb major* with three different expressive intentions: deadpan, projected, and exaggerated. Using a point-light technique (discussed in section 2.1.1) she continually tracked the location of the performer's head/face, collar, and ears, showing that the greatest inter-intention variability was observed in the head and collar, rather than the hands. Likely, this reflects the fact that hand movements are dictated by technical requirements, leaving little room for variation related to intended performance manner. Perceptual ratings of expressivity from another point-light piano performance suggest that although viewers are capable of using this visual information to infer performer intention regardless of the combination of points shown, head motion is the most salient type of movement differentiating intended performance manner.

#### 2.4.3 PITCH DISTANCE

Thompson and Russo (2007) have shown that the facial movements used by vocalists can be visually "read." They video-recorded three females singing intervals spanning 0 to 12 semitones, then measured the amount of head, mouth, and eyebrow movement, finding a strong correlation between movement amount and interval size. Furthermore, they showed that these differences are perceptible – the mean interval size judgments were well aligned with actual interval distance. As this study did not involve any auditory information, it is not clear from this experiment whether vision can actually influence the perception of heard interval size (although experiment 3 of Thompson et al., 2005 suggests this may be possible). Nonetheless, it demonstrates that vision is surprisingly effective at communicating intervallic information.

#### 2.4.4 LARGE-SCALE STRUCTURE AND NON-VERBAL COMMUNICATION

Other studies illustrating the communicative power of vision in musical contexts include the work of Krumhansl and Schenck (1997), showing striking parallels between the Minuetto from Mozart's *Divertimento No. 15*, and Balanchine's accompanying ballet choreography. Ratings of the structural and emotional properties shared strong similarities whether they were performed by participants hearing only the music, viewing only the dance, or both hearing the music and viewing the dance. In a different vein, case studies of the gestures and body positions used by Anne Lennox (Davidson, 2001) as well as the nonverbal stage behavior of *The Corrs* (Kurosawa & Davidson, 2005) show that visual information plays a role in a performer's interactions both with their audiences, and with one another.



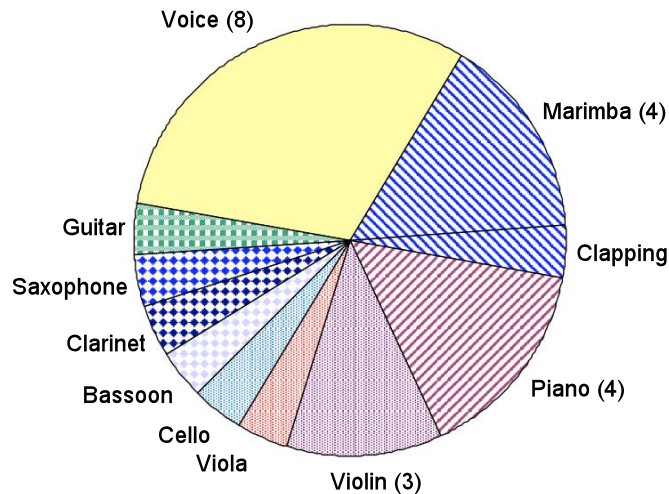
### 3 INSTRUMENTS AND INFLUENCE

One issue of importance in practically applying research on this topic is the degree to which documented visual influences are in fact instrument-specific. Can the types of influences invoked by violinists also be used by clarinetists and vocalists? Are any of these effects unique to specific instruments, and if so why?

#### 3.1 Which Instruments are Best at Using Visual Information?

Although research attention has shown a strong bias towards voice, percussion, and piano (Figure 7), in general the patterns of integration observed are neutral with respect to the instruments capable of employing them. Although some findings are for obvious reasons instrument-specific (e.g. that seeing lip movements increases lyric comprehensibility, Hidalgo-Barnes and Massaro 2007), these instances are likely exceptions rather than the rule. There is no a-priori reason to believe that the motions used by clarinetists to control musical tension (Vines et al., 2006) or the animated gestures used by marimbists to capture audience interest (Broughton & Stevens, in press) cannot be effectively employed by other instrumentalists and/or vocalists.

For example, that fact that vision is helpful in communicating expressive intentions has been shown using violinists (Davidson, 1993, experiment 1), pianists (Davidson, 1993, experiment 2), marimbists (Broughton & Stevens, in press), and clarinetists Vines et al. (2005). Yet the distribution of instruments used in this research suggests that vocalists, marimbists, and pianists may be particularly well-suited for effectively using visual information.



**Fig. 7.** Distribution of instruments used. Of the experiments summarized in Table 1, voice, percussion, piano, and violin were the most frequently employed [7]. The number in parentheses indicates the number of studies using each instrument (those without numbers appeared only once). Because of its acoustic properties, hand clapping is classified here as a percussive instrument.

##### 3.1.1 VOICE

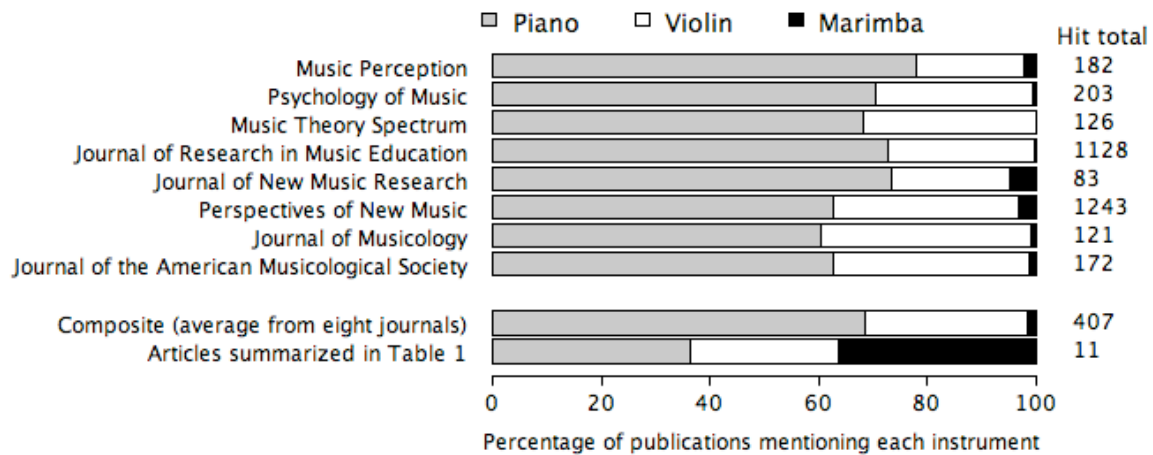
Given its clear favoritism (used in 31% [8] of the listings in Table 1), vocalists appear best able to effectively utilize visual information. As the eight instances of its use come from five different first-authors using a variety of methods, this bias is not merely reflective of the interests of one particular researcher or paradigm. Given the special resources recruited in the processing of speech, (Massaro & Stork, 1998; Middleweerd & Plomp, 1987; Sumbly & Pollack, 1954) and the privileged relationship between affective face and voice information (de Gelder, Pourtois & Weiskrantz, 2002), it is likely that this favoritism reflects in part patterns of integration unique to the voice. Furthermore, unrestricted by the need to hold an instrument, vocalists are able to move about onstage, using hand movements and/or facial expressions in a manner impossible for instrumentalists.

3.1.2 PERCUSSION

Aside from the voice, the most popular instruments used in this research were the piano, violin, and marimba. With a storied history in western classical music, it is not surprising that the first two have received a great deal of research attention. However the marimba is far less well-known, and rarely used in musical research.

A search of the full text of articles published in *Music Perception* revealed that although “piano” appears in 142 papers and “violin” in 36, “marimba” is mentioned in only 4. Of these four articles, one was Dahl and Friberg’s 2007 study on visual detection a performer’s emotional intentions! In a search of eight major musical journals, the term, “marimba” appeared in less than 2% of articles mentioning one of the three instruments, in contrast to 36% (4/11 studies) of the research on these three instruments summarized in Figure 7. Details of this search are given in Appendix I, and the results displayed graphically in Figure 8.

In comparison to the large number of articles contained in these journals, the scope of this review paper is much smaller. Although this makes drawing strong conclusions difficult, the juxtaposition of the marimba’s importance here with its obscurity in general raises the question of whether it too is in some way particularly well-suited for effectively using visual information.



**Fig. 8.** Specific mention of instruments in music research. The area of each horizontal bar reflects the number of articles within each journal mentioning a given instrument, relative to the total number of articles mentioning one of the three instruments. Numbers on the far right indicate the total hits for each journal. As shown in the bottom two bars, the marimba’s prominence in research on the role of visual information (bottom bar) far exceeds that which would be expected given its relative obscurity within the context of general musical scholarship (penultimate bar).

Given my interest and training in percussion, I cannot claim to be unbiased in my knowledge of the literature. However, I believe this survey is representative of research on this topic, and in fact reflects three factors making the marimba (and percussion in general) particularly adept at strategically using visual information:

- 1) Percussionists generally have less control over low-level aspects of music, such as note duration, than other musicians. This limitation provides the incentive for exploring techniques to compensate for such restrictions. For example, some expert percussionists have maintained that gestures can be used to overcome acoustic limitations regarding control over marimba note duration (Bailey, 1963).
- 2) Percussion is highly visual by nature, requiring more motion than is generally used by other instrumentalists. This in turn produces substantially more salient visual information than performances on non-percussive instruments such as the French horn or the flute.
- 3) Percussion instruments exhibit a clear causal relationship between gesture and sound, which is known to be a powerful cue triggering cross-modal integration (Schutz & Kubovy, in press; Schutz & Kubovy, 2008). The connection between individual notes and the movements used

to produce these notes is far less obvious for wind instruments, where changing pitches requires complex interactions between embouchure and fingerings that are far from transparent to audiences.

### 3.2 Other Observations

It is worth noting that most of the arguments for the special nature of percussion instruments such as the marimba apply (to some degree) to the piano as well. Given that the piano produces sound by striking a string with a hammer, it is in an acoustic sense a percussive instrument – and therefore suffers from some of the same limitations as the marimba (e.g. fixed tuning, limited control over note decay relative to wind and string instruments, etc). Playing it also requires a great deal of movement, movements that are clearly causally connected to the sounds produced. Therefore future research on the role of visual information in music may be most successful when focused on percussive instruments such as the marimba and piano (as well as the voice, for reasons discussed above in section 3.1.1).

Finally, although visual influences have been shown using vocalists as well as strings, woodwinds, and percussion, to the best of my knowledge there has not yet been any published research involving brass players. As there is no reason to believe they are any less capable of using visual information than other instruments, it is likely that this merely reflects a lack of research attention.

## 4 DISCUSSION

### 4.1 Relevance for Psychologists and Musicians

Research on this topic is of great importance to psychology, as it offers a unique vantage point from which to explore the mind's psychological structure. Given the ubiquity and antiquity of musical practice (Levitin, 2006), musically-based studies inherently capture many of the essential qualities lost in tightly controlled laboratory experiments. They can challenge (Schutz & Lipscomb, 2007) and even contribute to modifications (Schutz & Kubovy, in press) of generally accepted psychological theories by showing that artificially generated sounds do not always serve as accurate assessors of the perceptual system. Consequently, they serve as an invaluable supplement to traditional psychophysical research. Additionally, as music is of inherent interest to a wide population, research on its performance and perception will by default be relevant to a wide readership.

But ultimately, this topic is clearly of the greatest relevance to performing musicians (particularly soloists), whose careers rest in part on effective performer-audience communication. To this end, understanding the wide range of visual influences is crucial – from low-level, obligatory integration to high-level cognitive evaluations of performance quality. Of particular relevance is the fact that many of these manipulations are strategic, offering performers enhanced control over a variety of important musical properties. Some even overcome acoustic limitations – offering better comprehension of song lyrics (Hidalgo-Barnes & Massaro, 2007), and control over note duration (Schutz & Lipscomb, 2007).

### 4.2 The Musical Experience

Despite the indisputable role of visual information, it is important to keep in mind that music is an *auditory* phenomenon – even though it is clearly not a purely acoustic one. Ives' tongue-in-cheek question as to the role of sound in music aside, visual information is meaningful only in so far as it affects the way in which we perceive and/or evaluate sound. Visual information that fails to do so (or that affects the listening experience in an undesirable manner) plays no musical role, and in some cases may actually be harmful by distracting auditory attention (Schutz & Kubovy, 2006).

#### 4.2.1 ACOUSTICS AND MUSIC: AN IMPORTANT DISTINCTION

The fact that sound alone fails to describe the full musical experience raises the provocative question of *what music actually is*. The observations that (1) the influences discussed here were obtained using both non-musicians and professional musicians as evaluators and (2) some influences simply cannot be ignored lead to a somewhat heretical question – does acoustic information serve as an adequate description of the

musical experience? Are contexts devoid of visual information (e.g. CDs, radio broadcasts, iPods) presenting the full musical story, or are they robbing performers and listeners of an important dimension of musical communication? On a pragmatic level, do the blind auditions often used in selecting members of professional orchestras serve as a fair means of evaluation, given that such situations are not representative of the contexts in which much of the orchestra's playing actually takes place?

#### 4.2.2 VISION: AUGMENTING OR INTERFERING?

The larger question is whether visual information represents a distraction from "true" music (in which case it is problematic at best), or whether music is actually by definition truly multi-modal. Philosophical musings aside, practically speaking this question leads to a crossroad of sorts for performers and educators – should vision's influence be embraced, included as part of the standard musical curricula and treated as an essential component of music? Or should its role be downplayed, minimized whenever possible so as to avoid distracting from the "real" musical experience? To some, effort spent choreographing gestures and movements may seem like time wasted, time that could have been better spent working on phrasing and interpretation. Yet I believe there are two compelling reasons to embrace visual information as an important component of the musical experience.

The first is practical if not elegant. Simply put, given the multitude of research demonstrating that the brain is not ignorant of vision when processing sound, performers should not be either. It makes no more sense to ignore this fact on the grounds that it may be undesirable than it does to leave one's car doors unlocked because it would be nice to live in a world without crime. As vision clearly plays a role in the processing of sound, it is important for musicians to be aware of this fact, welcome or otherwise. The benefit of this understanding is the realization that vision is actually a helpful tool, capable of improving the quality of performer-audience interactions. Music is less about sound per-se than about using sound to create a particular experience within the mind of the listener. To this end, the strategic use of visual information is no less important than manipulations of breath control, bow position, striking angle, intonation, etc., and music will be only enhanced by appropriately embracing its communicative power.

The second (and more intriguing) reason for embracing vision's role is the realization that in many cases its influence on auditory perception is not a flaw – but rather an elegant feature. The brain is constantly engaged in a search for identity cues, probing whether auditory and visual information are derived from the same event (Spence et al., 2007; Vatakis & Spence, 2007; Vroomen, 1999; Welch, 1999; Welch & Warren, 1980). For when they are it is advantageous to integrate cross-modally, as multi-modal representations are in fact more accurate (Ernst & Banks, 2002). Consequently, our understanding of a speaker's message is improved when their lip movements and hand gestures affect a listener's perception of their spoken utterances. Likewise, our understanding of a performer's musical message is clearer and deeper when their gestures and facial expressions influence the audience's processing of their acoustic output. As such principles of auditory-visual integration were in place long before man's invention of music, the counter-intuitive nature of vision's role is indicative less of shortcomings in the design of the perceptual system than inaccuracies in the nature of our intuitions. In other words, given that vision has always affected human auditory perception, it has by definition been playing a role in the musical experience as long as there has been music to experience.

### 4.3 Conclusions

Ultimately, visual information does more than just "influence our perception and evaluation of sound" – it is actually an essential part of the music itself. It is of course entirely possible for music to exist in some form even in the absence of visual information. CDs and radio broadcasts capture a large percentage of the musical experience, and their lack of visual information no more precludes their enjoyment than the lack of visual information in a phone conversation prohibits spoken communication. And yet, most would agree that conversations held face-to-face are generally more fruitful and effective than those held over the phone, even though they involve essentially the same auditory information. This is because body language is an important component of spoken dialogue, resolving verbal ambiguities and offering useful information regarding a speaker's intentions and emotional state. So too it is with music – although it can occur in the absence of visual information, the experience is simply not the same.

Although some of the visual influences discussed here are quite large, many are smaller, representing "tweaks" to the perception of auditory information rather than large categorical shifts. To this

end, it is important to note that when distinguishing the merely good from the truly great, it is generally the small details that matter most. The acoustic differences with respect to note timing and volume (e.g. “shaping a line”) between dull and engaging performances are small – yet they give rise to vastly different musical experiences. Similarly, even small visual influences on the processing of sound can have important consequences, provided that they are employed in the proper manner.

For until the relatively recent advent of recording devices music had always been an audio-visual event, and the idea of experiencing one without the other would surely have seemed bizarre and foreign to our musical precursors. Indeed, Schuman once remarked of a well-known pianist that “he must be heard and seen; for if Liszt played behind a screen, a great deal of poetry would be lost.” (Mortgenstern, 1956, p. 155; as quoted by Davidson, 1995). Given the wide-spread availability of technology capable of supporting audio-visual transmission of music (DVDs, Youtube, iPhones, etc), perhaps in the 21<sup>st</sup> century it is time to remove our self-imposed screen, embrace visual information as a useful tool, and thereby regain the full musical experience of our ancestors.

## NOTES

[1] Please address all correspondence to Michael Schutz at [schutz@virginia.edu](mailto:schutz@virginia.edu). Mailing address: Department of Psychology, 102 Gilmer Hall, Charlottesville, VA 22904

[2] For evidence, see any live performance by the contemporary ensemble Eighth Blackbird. ([www.eighthblackbird.com](http://www.eighthblackbird.com)).

[3] These distinctions are primarily organizational, and reflect the clarity of the relationship between musical (perceptual) properties and their acoustic correlates. Properties such as pitch and duration (that are for the most part specified by acoustic information) are regarded as “low-level,” whereas emergent properties such as performance quality and emotional engagement (dependent upon the interaction of multiple acoustic parameters) are regarded as “high-level.” Properties not clearly falling into one of these categories are classified simply as “intermediate-level.”

[4] Psychological level of influence is classified according to the lowest level of influence demonstrated. For example, the results of Thompson et al. (in press) showing a visual influence on judgments of affect is classified on the perceptual level, even though this facial information can also be used communicatively in the absence of auditory information as shown by Thompson and Russo (2007). I have deliberately taken a “conservative” approach by defaulting to a cognitive categorization unless the perceptual criteria were clearly met. Studies exploring visual communication, rather than visual influence are regarded separately, as it is not yet clear whether they affect either the perception or cognition of sound.

[5] In this review, Thompson et al. (2005) discussed 5 experiments of relevance. Experiments 2 and 4 were subsequently expanded and published separately as Thompson and Russo (2007) and Thompson et al. (in press). For the sake of efficiency, I have therefore chosen to discuss these more detailed treatments (rather than experiments 2 and 4 from the 2005 review) throughout this paper.

[6] Performance quality ratings of females made in the audio-alone condition did not differ significantly between those females independently rated differently on attractiveness. Therefore, Wapnick et al. (1997) were not able to comment on whether attractiveness similarly influenced ratings of female singers. It should be noted that subsequent studies of this nature have also been confounded by a lack of variability in the audio-alone ratings.

[7] Classification follows that used in Table 1, in that a stimuli set used repeatedly (or with only small variations) within a single paper is counted as one instance, whereas different stimuli sets using different instruments within a single experiment are treated separately. For example, the two experiments discussed by Thompson et al. (in press) use the same recordings of faces and voices, and are therefore counted only once as they show the same type of visual influence. However, the two experiments discussed by Dahl and Friberg (2007) use stimuli depicting different performers on different instruments (marimba, bassoon, saxophone). Consequently, it appears three times in both the Table 1 and Figure 7, as it reflects visual

influences from three different instrumental sources. Davidson (1993, experiment 2) and Davidson (1994, experiment 2) were based on the same stimuli, and therefore are also represented only once in Figure 7.

[8] Of the 29 listings in Table 1, only 26 were used in calculating this percentage. Broughton and Stevens (in press) appears twice in the table, as its results discussing expressive intention and audience interest are listed separately. The findings of Krumhansl and Schenk (1997) were based on a ballet performance involving a symphony orchestra, and are therefore not relevant to an analysis of the distribution of the specific instruments used in this research. Finally, as discussed in endnote [7], Davidson (1993, experiment 2) and Davidson (1994, experiment 2) were based on the same stimuli, and therefore counted only once.

## REFERENCES

- Abry, C., Cathiard, M. A., Robert-Ribes J., & Schwartz J. L. (1994). The coherence of speech in audio-visual integration. *Current Psychology of Cognition*, Vol. 13, 52-59.
- Alais D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, Vol. 14, 257-262.
- Arieh, Y., & Marks, L. E. (2008). Cross-modal interaction between vision and hearing: A speed-accuracy analysis. *Perception & Psychophysics*, Vol. 70, No. 3, 412–421.
- Bailey, B. (1963). *Mental and Manual Calisthenics for the Mallet Player*. New York: Adler.
- Broughton, M., & Stevens, K. (in press). Music, movement, and the marimba: An investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychology of Music*.
- Byrne, D., London, O., & Reeves, K. (1968). The effects of physical attractiveness, sex and attitude similarity on interpersonal attraction. *Journal of Personality*, Vol. 36, 259-271.
- Calvert, G. A., Brammer, M. J., & Iversen S. D., (1998). Crossmodal identification *Trends in Cognitive Sciences*, Vol. 2, 247-253.
- Chapados, C., & Levitin, D. (in press). Cross modal interactions in the experience of musical performances: Physiological correlates. *Cognition*, doi:10.1016/j.cognition.2008.05.008.
- Clifford, M., & Walster, E. (1973). The effect of physical attractiveness on teacher expectation. *Sociology of Education*, Vol. 46, 248-258.
- Dahl, S., & Friberg, A. (2007). Visual Perception of Expressiveness in Musicians' Body Movements. *Music Perception*, Vol. 24, No. 5, 433-454.
- Dahl, S. (2000). The playing of an accent – preliminary observations from temporal and kinematic analysis of percussionists. *Journal of New Music Research*, Vol. 29, 225-233.
- Dahl, S. (2004). Playing the accent - comparing striking velocity and timing in an ostinato rhythm performed by four drummers. *Acta Acustica united with Acustica*, 90, 762-776.
- Davidson, J. W. (1993). Visual perception and performance manner in the movements of solo musicians. *Psychology of Music*, Vol. 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*, Vol. 26, No. 6, 279–301.
- Davidson, J. W. (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.

- Davidson, J. W. (2001). The role of the body in the production and perception of solo vocal performance: A case study of Annie Lennox. *Musicae Scientiae*, Vol. 5, 235–256.
- de Gelder, B., Pourtois, G., & Weiskrantz, L. (2002). Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proceedings of the National Academy of the Sciences*, Vol. 99, 4121-4126
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, Vol. 24, 285-290.
- Ekholm, E. (1994). *The effect of guided listening on evaluation of solo focal performance*. Master's thesis, McGill University.
- Ernst, M. O., & Banks M. S. (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, Vol. 415, 429-433.
- Finnäs, L. (2001). Presenting music live, audio-visually, or aurally – does it affect listeners' experiences differently? *British Journal of Music Education*, Vol. 18, No. 1, 55-78
- Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to ‘happy/ sad’ judgements in equitone melodies. *Cognition and Emotion*, 17, 25-40.
- Gillespie R. (1997). Ratings of violin and viola vibrato performance in audio-only and audiovisual presentations. *Journal of Research in Music Education*, Vol. 45, 212-220.
- Griffiths, N. & Davidson, J. (2006). The effects of concert dress and physical appearance on perceptions of female solo performers. In: M. Baroni, A. R. Addressi, R. Caterina, M. Costa (Eds) *Proceedings of the 9th International Conference on Music Perception & Cognition*, Bologna, Italy: pp. 1723-1726.
- Hidalgo-Barnes, M., & Massaro, D. (2007). Read my lips: An animated face helps communicate musical lyrics. *Psychomusicology*, Vol. 19, No. 2, 3-12.
- Huron, D. (2001) Is Music an Evolutionary Adaptation? *Annals of the New York Academy of Sciences*, Vol. 930, No. 1, 43–61.
- Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.
- Ives, Charles (1961). *Essays Before a Sonata*. New York: Norton.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, Vol 14, No. 2, 201-211.
- Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception*, Vol. 8, 189-202.
- King A. J., & Calvert G. A. (2001) Multisensory integration: Perceptual grouping by eye and ear. *Current Biology*, Vol. 11, 322-325.
- Krumnhansl, C., & Schenck, D. L. (1997). Can dance reflect the structural and expressive qualities of music? A perceptual experiment on Balanchine's choreography of Mozart's Divertimento No. 15. *Musicae Scientiae*, Vol. 1, No. 1, 63-85.
- Kurosawa, K., & Davidson, J. W. (2005). Nonverbal behaviours in popular music performance: A case study of The Corrs. *Musica Scientiae*, Vol. 9, No. 1, 111–133.

- Lapitsky, M., & Smith, C. M. (1981). Impact of clothing on impression of personal characteristics and writing ability, *Home Economics Research Journal*, Vol. 3, 98-102.
- Levitin, D. (2006). *This Is Your Brain on Music: The science of an obsession*. New York: Penguin Group.
- Massaro, D. W., & Stork, D. G. (1998). Speech recognition and sensory integration. *American Scientist*, Vol. 86, 236-244.
- McClaren, C. (1988). The Visual Aspect of Solo Marimba Performance. *Percussive Notes*, Vol. 26, No. 1, 54-58.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, Vol. 264, 746–748.
- Middleweerd, M. J., & Plomp, R. (1987). The effect of speechreading on the speech-reception threshold of sentences in noise. *Journal of the Acoustical Society of America*, Vol. 82, 2145-2146.
- Mortgenstern, S. (1956). *Composers on Music*. London: Faber & Faber.
- Primus (1993). *The Pressman*. From the Album *Pork Soda*. Santa Monica, CA: Interscope Records.
- Repp, B. H., & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological Research*, Vol. 68, 252-270.
- Rosenblum, L. D., & Fowler, C. A. (1991). Audiovisual investigation of the loudness-effort effect for speech and nonspeech events. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 17, 976-985.
- Saldaña H. M., & Rosenblum L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception & Psychophysics*, Vol. 54, 406-416.
- Saoud, E. (2003). The effect of stroke type on the tone production of the marimba. *Percussive Notes*, Vol. 41, No. 3, 40-46.
- Schutz, M., & Kubovy, M. (2006). Visual Gestures: Perceptual costs and benefits in the performance of live music. In: M. Baroni, A. R. Addressi, R. Caterina, M. Costa (Eds) *Proceedings of the 9th International Conference on Music Perception & Cognition*, Bologna, Italy: pp. 1551.
- Schutz, M., & Kubovy, M. (in press). Causality and audio-visual integration. *Journal of Experimental Psychology: Human Perception and Performance*.
- Schutz, M., & Kubovy, M. (2008, July). The effect of tone envelope on sensory integration: support for the 'unity assumption' (A). *Journal of the Acoustical Society of America*, Vol. 123, 3412.
- Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, Vol. 36, 888-897.
- Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, Vol. 145, 1328–1330.
- Spence, C., Sanabria, D., & Soto-Faraco, S. (2007). Intersensory Gestalten and crossmodal scene perception. In K. Noguchi (Ed.), *Psychology of beauty and Kansei: New horizons of Gestalt perception* (pp. 519–579). Tokyo, Japan: Nihon University College of Humanities and Sciences (distributed by Fuzanbo International).



Sumby, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, Vol. 26, 212-215.

Thompson, W. F., Russo, F. A., & Quinto, L. (in press). Audio-visual integration of emotional cues in song. *Cognition & Emotion*

Thompson, W. F., & Russo, F. A. (2007). Facing the Music. *Psychological Science*, Vol. 18, No. 9, 756-757.

Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica*, Vol. 156, No. 1, 203-227.

Timmers, R. Marlot, M., Camurri, M., & Volpe, G. (2006). Listener's emotional engagement with performances of a Scriabin étude: an explorative case study. *Psychology of Music*, Vol. 34, No. 4., 481-510.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics*, Vol. 69, No. 5, 744-756.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2005). Dimensions of Emotion in Expressive Musical Performance. *Annals of the New York Academy of Sciences*, Vol. 1060, pp. 462-466.

Vines, B., Krumhansl, C., Wanderley, M., & Levitin, D. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, Vol. 101, 80-113.

Vroomen, J. (1999). Ventriloquism and the nature of the unity decision. Commentary on Welch. In G. Aschersleben, T. Bachmann, & J. Musseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events*. Amsterdam, Netherlands: Elsevier, pp. 389-393.

Walker, J.T., & Scott K. J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones, and gaps. *Journal of Experimental Psychology: Human Perception and Performance*. Vol. 7, 1327-1339.

Wanderley, M. M., & Vines, B. W. (2006). Origins and functions of clarinetists' ancillary gestures. In E. King & A. Gritten (Eds.), *Music and Gesture: New Perspectives on Theory and Contemporary Practice* Hampshire, UK: Ashgate, pp. 165-191.

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., & Hatch, W. (2005) The Musical Significance of Clarinetists' Ancillary Gestures: An Exploration of the Field. *Journal of New Music Research*, Vol. 34, No. 1, 97-113.

Wapnick, J., Ryan, C., Lacaille, N., & Darrow, A.-A. (2004). Effects of selected variables on musicians' ratings of high-level piano performances. *International Journal of Music Education*, Vol. 22, No. 1, 7-20.

Wapnick, J., Darrow, A.-A., Kovacs, J., & Dalrymple, L. (1997). Effects of physical attractiveness on evaluation of vocal performance. *Journal of Research in Music Education*, Vol. 45, No. 3, 270-479.

Wapnick, J., Kovacs Mazza, J., & Darrow, A.-A. (1998). Effects of performer attractiveness, stage behavior, and dress on vocal performance evaluation. *Journal of Research in Music Education*, Vol. 46, No. 5, 510-521.

Wapnick, J., Mazza, J. K., & Darrow, A.-A. (2000). Effects of Performer Attractiveness, Stage Behaviour and Dress on Evaluations of Children's Piano Performances, *Journal of Research in Music Education*, Vol. 48, No. 4, 323-336

Welch, R. B. (1999). Meaning, attention, and the 'unity assumption' in the intersensory bias of spatial and

temporal perceptions. In G. Aschersleben, T. Bachmann, & J. Musseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events*. Amsterdam, Netherlands: Elsevier, pp. 371–387.

Welch, R. B., & Warren D. H. (1980). Immediate response to intersensory discrepancy. *Psychological Bulletin*, Vol. 88, 638-667.

## APPENDIX I: JOURNAL SURVEY

To gauge the relative focus of music research on the piano, violin, and marimba, a variety of journals were surveyed using online tools available on their websites. All queries were made on Aug 1, 2008, using the full range of search dates available. The earliest year of back indexing available is listed in the table below. For cases in which all years are indexed, this number indicates the year of its inaugural publication (marked with a \* symbol). Websites used to conduct the searches are indicated below Table 2.

**Table 2.** Results from instrument queries by journal. Second line includes earliest year indexed and available for online searching at the time this search was conducted.

	Music Perception	Psychology of Music	Music Theory Spectrum	Journal of Research in Music Education	Journal of New Music Research	Perspectives of New Music	Journal of Musicology	Journal of the American Musicological Society
<i>Years covered in search</i>	2000-present	1973*-present	2000-present	1953-2006	1972*-present	1962-2000	2000-present	2000-present
Piano	142	143	86	819	61	780	73	108
Violin	36	59	40	305	18	421	47	62
Marimba	4	1	0	4	4	42	1	2
<b>Total for all 3 instruments</b>	<b>182</b>	<b>203</b>	<b>126</b>	<b>1128</b>	<b>83</b>	<b>172</b>	<b>121</b>	<b>172</b>

### Websites used for the searches:

Music Perception <http://caliber.ucpress.net/search/advanced>

Psychology of Music <http://pom.sagepub.com/>

Music Theory Spectrum <http://caliber.ucpress.net/search/advanced>

Journal of Research in Music Education <http://www.jstor.org/action/showAdvancedSearch>

Journal of New Music Research <http://www.informaworld.com/smpp/title~content=t713817838>

Perspectives of New Music <http://www.jstor.org/action/showAdvancedSearch>

Journal of Musicology <http://caliber.ucpress.net/search/advanced>

Journal of the American Musicological Society <http://caliber.ucpress.net/search/advanced>

## APPENDIX II: ONLINE SAMPLE STIMULI

Links to samples of stimuli used in some experiments can be found online at:

McGurk and MacDonald (1976) – [http://www.media.uio.no/personer/arntm/McGurk\\_english.html](http://www.media.uio.no/personer/arntm/McGurk_english.html)

Schutz and Lipscomb (2007) – <http://www.acoustics.org/press/153rd/schutz.html>

Hidalgo-Barnes & Massaro (2007) – <http://mambo.ucsc.edu/psl/primus.html>

Additionally, the following companion site for this article will be periodically updated to include additional relevant stimuli: <http://people.virginia.edu/~mrs8q/seeingmusic.html>