
Hearing gestures, seeing music: Vision influences perceived tone duration

Michael Schutz[¶], Scott Lipscomb

School of Music, Northwestern University, Evanston, IL 60208, USA

Received 25 May 2006, in revised form 17 November 2006; published online 8 June 2007

Abstract. Percussionists inadvertently use visual information to strategically manipulate audience perception of note duration. Videos of long (L) and short (S) notes performed by a world-renowned percussionist were separated into visual (Lv, Sv) and auditory (La, Sa) components. Visual components contained only the gesture used to perform the note, auditory components the acoustic note itself. Audio and visual components were then crossed to create realistic musical stimuli. Participants were informed of the mismatch, and asked to rate note duration of these audio-visual pairs based on sound alone. Ratings varied based on visual (Lv versus Sv), but not auditory (La versus Sa) components. Therefore while longer gestures do not make longer notes, longer gestures make longer sounding notes through the integration of sensory information. This finding contradicts previous research showing that audition dominates temporal tasks such as duration judgment.

1 Introduction

Perception of everyday events involves integration of information delivered through multiple senses. Resolution of conflicting sensory information depends on the type of task; generally vision governs spatial awareness, whereas audition is favored for temporal judgments such as rating tone length. Exceptions to this pattern are rare and observed only when the typically dominant mode is ambiguous; therefore deviations using unambiguous information would challenge current explanations of sensory integration. Inspired by an ongoing debate among musicians, we used real-world audio and visual information to demonstrate a natural auditory illusion reversing typical dominance patterns.

The marimba is played by raising a mallet into the air before lowering it rapidly and ‘bouncing’ it off the bar. As a result of this impact, energy from the mallet is translated into sound waves creating a musical note. Percussionists agree that the sound of this note is a function of the angle of attack, the force with which the mallet strikes the bar, how the mallet is held, and the type of mallet used. However, they routinely disagree whether the gesture length (eg distance covered by the ‘up–down’ motion) has any direct effect on the length of the resulting note.

Some believe longer gestures create longer notes (Bailey 1963). Others insist gesture length cannot alter note duration independently of intensity (Stevens 1990), an opinion bolstered by failure to find acoustic distinction between notes performed with different gestures (Saoud 2003). Both views initially appear quite reasonable—much as a longer swing of the bat sends the ball farther, it is plausible longer gestures produce longer notes. On the other hand, if energy is transferred from mallet to bar according to $e = \frac{1}{2}mv^2$, differences in gesture length are irrelevant, as velocity and mass fully dictate the physics of the impact. Such reasoning led renowned marimbist Leigh Howard Stevens (2004) to conclude: “stroke height has no more to do with [note] duration than the sound of a car crashing is dependent on how long a road trip was taken before the accident”.

[¶] Current and mailing address: Department of Psychology, University of Virginia, 102 Gilmer Hall, Charlottesville, VA 22904-4400, USA; e-mail: schutz@virginia.edu

While perceptual psychologists are well aware of the disjunct between physical properties and the perceptual experience of those properties, there is no evidence in the sensory-integration literature to suggest vision is capable of altering auditory-duration judgments. This study was designed to clarify the role of gesture in music performances by examining the individual contributions of auditory and visual information on audience perception of note length. Finding a visual influence on perceived note length would break from previous sensory-integration research by demonstrating vision can influence auditory-duration judgments, which would have significant implications for theories of sensory integration. It would also demonstrate a useful technique employed by expert musicians to strategically control the musical experience, which would have implications for our understanding of the performance and perception of music.

1.1 *Sensory integration*

The integration of information from multiple sensory inputs is crucial to functioning in a multi-sensory world, and there has been much research into both the experiential (Shimojo and Shams 2001) and neurophysiological (Calvert et al 1998; King and Calvert 2001) aspects of sensory integration. Patterns of influence are generally a function of task type. The ‘ventriloquist effect’, in which speech appears to originate from the moving lips of an inarticulate puppet, is a classic example of vision ‘capturing’ (eg controlling) auditory localization (Jack and Thurlow 1973). While there has been much interest in the integration of audio and visual speech information (Abry et al 1994), for purposes of this review we have chosen to focus on non-speech literature owing to clear differences in the processing of speech and non-speech sounds. Non-speech parallels of the ventriloquist illusion presenting visual and auditory stimuli originating from different locations demonstrate people experience both as emanating from the location of the visual information (Thomas 1941; Witkin et al 1952; Jackson 1953; Bertelson and Radau 1981; Bertelson et al 2000). Because of its superior spatial acuity, vision dominates spatial tasks.

Owing to superior temporal precision, audition generally governs temporal tasks. When presented with simultaneous flash/tone pairs and asked to rate flash length independently of tone duration, ratings of the flash component of the flash/tone pairs were similar to ratings of the tone component presented in isolation (Walker and Scott 1981). Auditory primacy has also been reported when counting the number of visual flashes paired with tones (Shams et al 2002), matching the rate of visual flicker with auditory flutter (Shipley 1964; Welch et al 1986), and estimating temporal order of flash and tone pairings (Fendrich and Corballis 2001). The consistent observation of auditory dominance in temporal tasks along with visual dominance in spatial tasks led to formulation of the ‘modality-appropriateness hypothesis’ (Welch and Warren 1980), stating modality dominance is determined by the type of task measured (eg spatial, temporal). Given high quality (eg unambiguous) information, this is a good rule-of-thumb, as exceptions occur only when the typically dominant mode is ambiguous.

Wada et al (2003) demonstrated one such exception, pairing auditory tones increasing, decreasing, or constant in their rate of flutter with visual lights increasing, decreasing, and constant in their rate of flicker. Participants were asked to identify the change in presentation rate (eg increasing versus decreasing) independently for each modality. As predicted by the modality-appropriateness hypothesis, visual flicker rate was irrelevant to estimations of auditory flutter when flutter was increasing or decreasing. However, it did influence estimations when the flutter rate was ambiguous (eg constant, offering no information regarding rate change). Other work has shown that decreasing the quality of information in the stronger modality can temper typical dominance patterns (Battaglia et al 2003), even to the point of reversal in extreme cases (Alais and Burr 2004).

Together, these results concur with the ‘optimal-integration hypothesis’ which states that the direction of influence is based on information quality rather than information modality (Ernst and Banks 2002). While several studies have successfully shown a reversal of typical modality-based dominance patterns, there is no evidence to suggest dominance pattern reversal with unambiguous stimuli.

Both the optimal-integration and modality-appropriateness hypotheses agree that audition is superior at detecting temporal changes, and vision should not influence duration judgments of clearly differentiable auditory tones (Walker and Scott 1981). However, the belief among some percussionists that longer gestures produce longer notes persists despite strong opinions (Stevens 2004) and acoustic evidence (Saoud 2003) to the contrary, hinting vision may in fact play a role in the perception of note length. The purpose of the present study is to determine to what extent (if any) the visual gesture influences audience perception of note duration. Discovery of a visual influence on perceived note duration would demonstrate marimbists display a tacit understanding of this auditory illusion and offer satisfying resolution to the debate over whether gesture length influences note duration. Furthermore, it would provide novel evidence of vision influencing a temporal task involving unambiguous auditory information, which would require refining current theories of sensory integration.

2 Experiment

2.1 Research questions

We investigate the relationship between performance gesture and perceived note duration through the following questions:

1. Does gesture length affect acoustic note length?
2. Does gesture length affect perceptual note length under the conditions:
 - (a) audio alone—without visual gesture information;
 - (b) audio-visual—with visual gesture information?

2.2 Method

2.2.1 Participants. Fifty-nine Northwestern University undergraduate music majors between the ages of 18 and 23 years participated in return for extra credit in their music theory or aural skills classes. While participants were all trained musicians, none considered percussion their primary instrument.

2.2.2 Stimuli. Marimba virtuoso Michael Burritt performed a series of notes on a Malletech Roadster marimba in the Master Class Room of Northwestern University’s Regenstein Hall, using Malletech MB-8 and MB-13 mallets. The pitch levels chosen were evenly spread across the instrument: E1 (~82 Hz), D4 (~587 Hz), and G5 (~1568 Hz). The performer held four mallets as if performing in an actual recital and was visible from the waist upwards, capturing his full range of motion. A Cannon model GL1 camera was used to record video, and Audio-Technica AT4041 microphones for audio.

The marimbist performed two stroke types—‘long’ and ‘short’. Additionally, he performed a ‘damped’ note by muffling the bar with his free hand immediately after striking to produce an artificially shortened note. As shown in table 1, the ‘long’ (L) and ‘short’ (S) stroke types were subsequently separated into auditory (La, Sa) and visual (Lv, Sv) components. The auditory component from a damped note was also extracted, yielding three auditory (La, Sa, Da) and two visual (Lv, Sv) stroke types which were combined factorially to create the six audio-visual stimuli used for each pitch level in the audio-visual condition. The three auditory components were also presented without visual information in the audio-alone condition.

Sa, La, and Da were peak-normalized within each pitch level, preserving their relative relationships while highlighting potential differences in duration. Stimuli were

Table 1. Stroke types broken into auditory and visual components (only the auditory portion of the damped stroke type was used). This results in six pairings for each pitch level.

Original stroke type	Auditory component	Visual component
Long	La	Lv
Short	Sa	Sv
Damped	Da	



Figure 1. Stimuli showed the upper body of the marimbist, including full stroke preparation and release.

rendered in AVI format with Sonic Vegas Pro video editor, then converted to MPEG files with a Cinema Craft encoder. Care was taken to ensure each video was at the proper angle and depth to show full stroke preparation and release (see screen shot in figure 1).

2.2.3 Procedure. Participants were informed at the outset that some of the stimuli contained auditory and visual components that had been intentionally mismatched. The experiment took place in a quiet room at the Northwestern University Library; Dell Dimensions 4100 computers were used with Dell UltraSharp model 1800FP 18.1 inch monitors equipped with Sony MDR-7506 Dynamic Stereo headphones. The stimuli were presented in blocks organized into two conditions: (i) as audio-visual stimuli combining the visual gesture and auditory note, and (ii) as audio-alone. The audio-visual condition always preceded audio-alone. Participants were allowed to adjust the playback volume as desired during the warm-up period.

Participants indicated perceived duration using an unmarked, 101 point slider with endpoints labeled 'short' and 'long'. Before starting the experiment they were instructed to base their duration rating in the audio-visual condition on the auditory information alone. However, to ensure they were attending to the visual information, participants were also asked to respond to a second question concerning the level of agreement between the visual and auditory components of the stroke. Responses were made on a second, unmarked slider with endpoints labeled 'minimum agreement' and 'maximum agreement'. The primary purpose of the discrepancy rating was to force participants to attend to information in both modalities. Previous experiments have demonstrated that asking participants to provide discrepancy ratings does not interfere with the primary task of judging the information itself (Rosenblum and Fowler 1991; Saldaña and Rosenblum 1993).

2.3 Results

The results are summarized in three sections, paralleling the sub-questions regarding the effect of gesture on acoustic and perceptual note length. Acoustic duration was analyzed by selecting arbitrary cutoff values, then measuring the time at which a given note dropped below the defined threshold. Perceptual-duration ratings for each stimulus were collapsed across trials, then analyzed with separate ANOVAs for the audio-alone and audio-visual conditions. Finally, analysis of the collapsed agreement ratings for the audio-visual condition showed participants were sensitive to the pairing of audio and visual information. In all analyses a threshold of $p < 0.05$ was used to assess statistical significance.

2.3.1 Acoustic length. The acoustic profile shown in figure 2 suggested clear differences between damped (Da) and undamped (La, Sa) stroke types within the lowest two pitch levels, as well as an overall distinction between each of the low, medium, and high pitch levels. However, on the basis of the graph of their profiles Sa and La appeared to be indistinguishable. This observation was confirmed by picking twenty-one cutoff points in the range of the logarithm of rms amplitude (-3 , -5), then measuring the time at which the acoustic profile of each stroke type first dropped below this threshold (mean cutoff times: Sa = 0.4594 s, La = 0.4424 s). After applying a square root transformation to correct for a skewed distribution (transformed cutoff times: Sa = 0.6294 s, La = 0.6268 s), a t -test revealed no difference ($t_{122.18} = 0.0604$, $p = 0.952$) between the cutoff times. Therefore, we conclude Sa and La were acoustically indistinguishable.

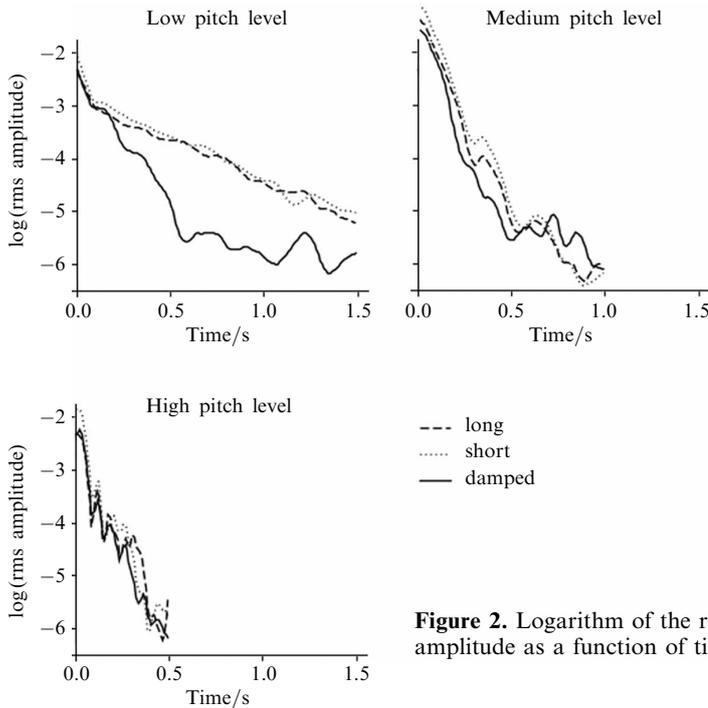


Figure 2. Logarithm of the rms (root mean square) of the amplitude as a function of time for each pitch level.

2.3.2 Perceptual length: Audio-alone. Duration ratings in the audio-alone condition were assessed with a 3 (pitch) \times 3 (auditory stroke type) repeated-measures ANOVA with both pitch and auditory stroke type as within-participants variables. The most important finding was that, while there was also a main effect of auditory stroke type ($F_{2,116} = 278.4$, $p < 0.0001$), a series of planned comparisons demonstrated that this merely reflected a distinction between damped and undamped notes rather than perceptible

differences between the short and long stroke types, as shown in table 2. There was also a main effect of pitch level ($F_{2,116} = 138.125$, $p < 0.0001$), reflecting that lower pitches ring for a longer amount of time. A two-way interaction was observed between auditory stroke type and pitch level ($F_{4,232} = 174.996$, $p < 0.0001$), suggesting the possibility of differences between stroke types within some pitch levels. However, as shown in figure 3 and table 2, participants were unable to consistently differentiate between Sa and La.

Table 2. Within-condition planned comparisons using Bonferroni corrections.

Comparison	Difference	95% Confidence	p Value
Audio-alone			
La–Sa	1.957	± 2.196	0.097
Sa–Da	27.712	± 3.888	< 0.0001
La–Da	25.759	± 3.375	< 0.0001
Audio–visual			
La–Sa	0.365	± 3.224	0.643
Sa–Da	20.150	± 1.568	< 0.0001
La–Da	19.784	± 1.568	< 0.0001
Lv–Sv	20.047	± 3.110	< 0.0001

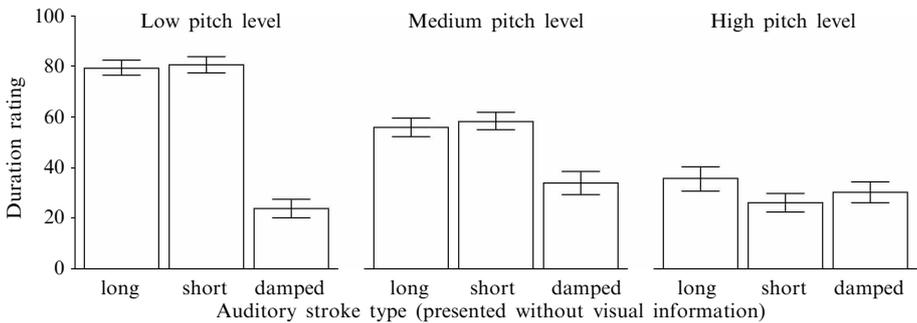


Figure 3. Duration ratings in the audio-alone condition for each of the three auditory components separated by pitch level.

Although slight differences were observed between Sa and La at the highest pitch level, they were small in size, did not appear in the audio-visual condition, never replicated in subsequent experiments, and were similar to those reported by Saoud (2003) among stroke types intended to be identical. Therefore we interpret this difference as a byproduct of using natural stimuli rather than a ‘true’ difference produced intentionally by the performer.

2.3.3 Perceptual length: Audio-visual. Duration ratings in the audio-visual condition were assessed with a 3 (pitch) \times 3 (auditory stroke type) \times 2 (visual stroke type) repeated-measures ANOVA with pitch, auditory stroke type, and visual stroke type as within-participants variables. The most important finding was a significant effect of visual stroke type ($F_{1,58} = 166.5$, $p < 0.0001$), indicating a strong visual influence (partial $\eta^2 = 0.742$) despite explicit instructions for participants to base responses on hearing alone (figure 4). The main effect of auditory stroke type ($F_{2,116} = 153.9$, $p < 0.0001$) was again due only to the distinction between damped and undamped stroke types (table 2). While there was also a main effect of pitch level ($F_{2,116} = 162.774$, $p < 0.0001$), no significant interaction was observed between auditory and visual stroke types ($F_{1,116} = 767$, $p = 0.467$).

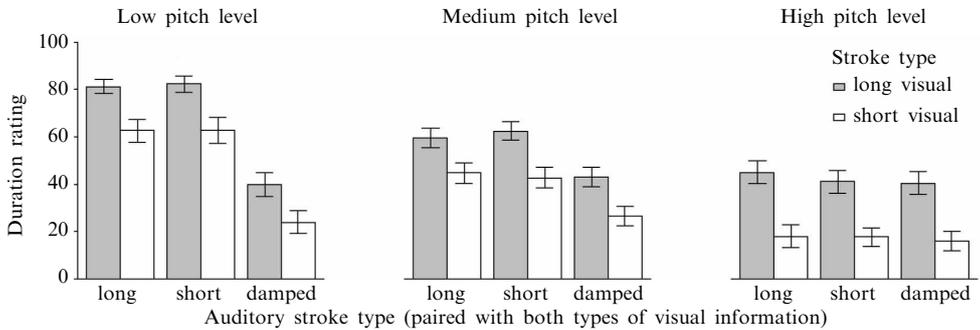


Figure 4. Duration ratings in the audio-visual condition for each of the three auditory stroke types separated by pitch level, when paired with long and short visual stroke types.

2.3.4 Agreement ratings. Agreement ratings were assessed with a 3 (pitch) \times 3 (auditory stroke type) \times 2 (visual stroke type) repeated-measures ANOVA with pitch, auditory stroke type, and visual stroke type as within-participants variables. This revealed a significant interaction between audio and visual stroke type ($F_{2,116} = 51.136, p < 0.0001$), indicating participants were attentive to the levels of audio-visual agreement. As the purpose of the agreement question was merely to ensure that participants were not ignoring the visual information, it is sufficient to note that participants were sensitive to the pairing of auditory and visual information.

3 Discussion

As shown by the quantitative analysis, there was no meaningful distinction between the ratings for Sa and La in either condition—therefore gesture length is irrelevant in the absence of visual information. However, owing to differences in ratings based on visual stroke type in the audio-visual condition we conclude that variations in gesture allow the performer to create short and long sounding notes, provided the audience is watching as well as listening.

As these results were replicated in a subsequent experiment with participants drawn from an introductory psychology course, they are not dependent upon formal musical training. Therefore our results document a real-life cross-modal interaction, demonstrating a previously unreported finding of vision influencing audition in a temporal task using unambiguous auditory information. This runs counter to previous results for tone length estimation in the presence of audio-visual discordance (Walker and Scott 1981) and the general pattern of auditory dominance in temporal tasks (Shipley 1964; Welch et al 1986; Fendrich and Corballis 2001; Shams et al 2002).

As the modality-appropriateness hypothesis predicts dominance of the modality with the greatest acuity in the domain of interest (Welch and Warren 1980), it is clearly incompatible with our results. The optimal-integration hypothesis could predict such results if information in the non-dominant modality were of higher quality owing to ambiguity in the generally dominant modality (Ernst and Banks 2002; Alais and Burr 2004). While Sa and La were ‘ambiguous’ *within* each pitch level, as shown in figure 2 there were clear differences in duration *between* pitch levels, as well as between damped and undamped notes at the lowest two pitch levels. These acoustic differences were reflected in perceptual ratings under both audio-alone (figure 3) and audio-visual (figure 4) conditions.

Our results show a reversal of traditional dominance patterns based on unambiguous information, and are therefore not compatible with either the optimal-integration or modality-appropriateness hypotheses. Further research is needed to understand which aspects of our stimuli drive this previously unobserved phenomenon.

There are three main differences between our stimuli and the flash/tone pairings generally studied which may contribute to our unusual results. First, previous studies generally used auditory information with clear onset/offset points, whereas many of the marimba notes in our stimuli decay gradually over time. However, since the visual influence was strong even for relatively short notes (eg damped or at the higher pitch level) this difference alone cannot account for the consistent visual influence across all auditory stroke types and pitch levels. Second, our stimuli show stroke preparation followed by an impact coinciding with note onset, making it clear that the gesture caused the note. Therefore unlike artificial flash/tone pairings, the audio and visual components of each marimba stroke are both representative of real-world information and share a meaningful relationship. Finally, our stroke-type gestures are dynamic, whereas visual stimuli in most sensory-integration research are static.

While previous research suggests that visual information should not play any role in temporal judgments of unambiguous auditory information, the nature of the relationship between information in each modality can be a factor governing integration (de Gelder et al 2002; Pourtois and de Gelder 2002). The ventriloquist illusion depends upon spatial agreement between the audio and visual information (Jack and Thurlow 1973) much as the degree of influence in the non-attended modality is mediated by the plausibility of its relationship to the attended modality (Watanabe and Shimojo 2001).

To the best of our knowledge, this study represents the only documentation of a natural audio-visual interaction in a non-speech task. Most sensory integration research involves artificial tones and light flashes which are not representative of the rich sensory information encountered in the everyday world. While built upon 'natural' audio and visual source material, the speech information used in the McGurk (McGurk and MacDonald 1976) effect, clapping sounds and images in Rosenblum and Fowler (1991), and cello bowing/plucking in Saldana and Rosenblum (1993) employ artificial pairings to show intriguing but unrealistic demonstrations of sensory integration. Other studies use realistic musical stimuli to show a role of visual information in music perception for judgments of 'musical expressivity' (Davidson 1993) and 'string vibrato quality' (Gillespie 1997). While stimulating and thought-provoking for musicians, the subjective nature of such dependent measures and impracticality of acoustic analysis make it difficult to draw strong conclusions about the degree of sensory integration demonstrated.

Our artificial pairings (Lv Sa, Sv La) served only to demonstrate the inconsequential nature of the auditory component of the stroke type. None of the stroke-type information in either modality was changed in any substantive way. The true long (Lv La) and short (Sv Sa) stroke types occur naturally in recitals whenever marimbists attempt to create long and short notes. While failing to create notes that are acoustically long and short, they are (accidentally) successful in creating long and short *sounding* notes. In using gesture information to strategically control perceived note length, some percussionists have inadvertently stumbled upon a practical application of sensory integration and hidden it in full view of concert audiences (and experimental psychologists) for centuries.

3.1 *Who was right?*

Our results concur with previous work reporting no acoustic distinction in note duration between marimba stroke types (Saoud 2003). Consequently, there were no differences between the stroke types when presented without visual gesture information, supporting the position articulated by Stevens (1990, 2004). However, the difference between true long (Lv La) and true short (Sv Sa) stroke types when presented with visual information validates Bailey's assertion that it is possible to produce short and long notes on the marimba. We conclude that the difference in duration between long and

short marimba notes is ‘perceptual’ rather than ‘real’, caused by visual artifacts of the performer’s acoustically inconsequential gesture. Resolution of the debate requires recognizing it was never the answers that disagreed, but rather the questions.

It is worth noting that the lack of distinction between Sa and La was not a result of our choice of performer, an internationally acclaimed solo marimbist and Professor of Percussion at Northwestern University. If he was unable to produce acoustic differences through manipulation of gesture length then it simply cannot be done. However, while unable to alter the sound of the note, unbeknownst to the performer, his gesture serendipitously alters the way the note sounds, thereby (accidentally) overcoming a profound limitation of the instrument.

That audiences distinguish note length through sensory integration rather than acoustic recognition is irrelevant from the musical perspective. Strategic use of gesture gives the performer a mechanism for shaping the audience’s musical experience. Musical communication relies on correlation not between performer intent and acoustic result, but between performer intent and audience perception. Skilled performers accomplish musical communication by sidestepping the impractical, ‘correcting’ faulty acoustic information to align audience experience with performer intention.

Our findings demonstrate contexts that ignore visual information (radio broadcasts, recorded performances, blind auditions, etc) are robbing both the performer and audience of a significant dimension of musical communication. Given the observed disjunct between sound and its perception, it is important to remember that music is only music within the mind of the listener. Virtuosos are masters at shaping the musical experience—which in this case means using visual information to accomplish that which is impossible ‘in reality’.

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* **13** 52–59
- Alais D, Burr D, 2004 “The ventriloquist effect results from near-optimal bimodal integration” *Current Biology* **14** 257–262
- Bailey B, 1963 *Mental and Manual Calisthenics for the Mallet Player* (New York: Adler)
- Battaglia P W, Jacobs R A, Aslin R N, 2003 “Bayesian integration of visual and auditory signals for spatial localization” *Optical Society of America* **20** 1391–1397
- Bertelson P, Radeau M, 1981 “Cross-modal bias and perceptual fusion with auditory–visual spatial discordance” *Perception & Psychophysics* **29** 578–584
- Bertelson P, Vroomen J, Gelder B de, Driver J, 2000 “The ventriloquist effect does not depend on the direction of deliberate visual attention” *Perception & Psychophysics* **62** 321–332
- Calvert G A, Brammer M J, Iversen S D, 1998 “Crossmodal identification” *Trends in Cognitive Sciences* **2** 247–253
- Davidson J W, 1993 “Visual perception of performance manner in the movements of solo musicians” *Psychology of Music* **21** 103–113
- Ernst M O, Banks M S, 2002 “Humans integrate visual and haptic information in a statistically optimal fashion” *Nature* **415** 429–433
- Fendrich R, Corballis P M, 2001 “The temporal cross-capture of audition and vision” *Perception & Psychophysics* **63** 719–725
- Gelder B de, 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 of the USA* **99** 4121–4126
- Gillespie R, 1997 “Ratings of violin and viola vibrato performance in audio-only and audiovisual presentations” *Journal of Research in Music Education* **45** 212–220
- Jack C E, Thurlow W R, 1973 “Effects of degree of visual association and angle of displacement on the ‘ventriloquism’ effect” *Perceptual & Motor Skills* **37** 967–979
- Jackson C, 1953 “Visual factors in auditory localization” *Quarterly Journal of Experimental Psychology* **5** 52–65
- King A J, Calvert G A, 2001 “Multisensory integration: Perceptual grouping by eye and ear” *Current Biology* **11** r322–r325
- McGurk H, MacDonald J, 1976 “Hearing lips and seeing voices” *Nature* **264** 746–748

-
- Pourtois G, Gelder B de, 2002 "Semantic factors influence multisensory pairing: a transcranial magnetic stimulation study", *Cognitive Neuroscience and Neuropsychology* *NeuroReport* **13** 1567–1573
- 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* **17** 976–985
- Saldana H M, Rosenblum L D, 1993 "Visual influences on auditory pluck and bow judgments" *Perception & Psychophysics* **54** 406–416
- Saoud E, 2003 "The effect of stroke type on the tone production of the marimba" *Percussive Notes* **41**(3) 40–46
- Shams L, Kamitani Y, Shimojo S, 2002 "Visual illusion induced by sound" *Cognitive Brain Research* **14** 147–152
- Shimojo S, Shams L, 2001 "Sensory modalities are not separate modalities: plasticity and interactions" *Current Opinion in Neurobiology* **11** 505–509
- Shipley T, 1964 "Auditory flutter-driving of visual flicker" *Science* **145** 1328–1330
- Stevens L H, 1990 *Method of Movement* 2nd edition (Ashbury Park, NJ: Keyboard Percussion Publications)
- Stevens L H, 2004, personal communication (e-mail)
- Thomas G, 1941 "Experimental study of the influence of vision on sound localization" *Journal of Experimental Psychology* **28** 163–175
- Wada Y, Kitagawa N, Noguchi K, 2003 "Audio-visual integration in temporal perception" *International Journal of Psychophysiology* **50**(1–2) 117–124
- 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* **7** 1327–1339
- Watanabe K, Shimojo S, 2001 "When sound affects vision: Effects of auditory grouping on visual motion perception" *Psychological Science* **12** 109–116
- Welch R B, DuttonHurt L D, Warren D H, 1986 "Contributions of audition and vision to temporal rate perception" *Perception & Psychophysics* **39** 294–300
- Welch R B, Warren D H, 1980 "Immediate response to intersensory discrepancy" *Psychological Bulletin* **88** 638–667
- Witkin H A, Wapner S, Leventhal T, 1952 "Sound localization with conflicting visual and auditory cues" *Journal of Experimental Psychology* **43** 58–67

Conditions of use. This article may be downloaded from the E&P website for personal research by members of subscribing organisations. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher.