# INFLUENCE OF VISUAL INFORMATION ON AUDITORY PERCEPTION OF MARIMBA STROKE TYPES

*Michael Schutz and Scott D. Lipscomb*[i]

Northwestern University

## ABSTRACT

While cross-modal interactions have been well explored in speech perception, there has been comparatively little research on musical contexts. Such an effect would be quite interesting as music is typically considered to be a primarily auditory experience. Percussionists must use a large amount of physical motion to produce sound. While differences in the physical gestures used to produce different stroke types is quite clear visually, the audible effect of these physical gestures is far more ambiguous. These factors make evaluation of stroke type (staccato, normal, legato) on percussion instruments a strong candidate for the study of the influence of visual information on auditory perception. The present study investigates the degree to which visual information affects the auditory perception of stroke type. Specifically, we are interested in the resolution of conflicting visual and auditory information, and the degree to which visual information from physical gestures made by the performer might influence listeners' auditory perception of stroke type.

## 1. BACKGROUND

Extensive research on cross-modal interactions has shown significant effects of visual information on auditory perception. In perhaps the most remarkable example, McGurk and MacDonald (1976) demonstrated that when presented with contrasting audio and visual speech sounds, many listeners believe they are hearing something other than the aurally presented phonemes. Psychophysical studies have shown that visual information can effect auditory perception in such non-speech tasks as sound localization (Bertelson & Radeau, 1977, 1981; Weerts & Thurlow, 1971), hand-clapping loudness evaluations (Rosenblum & Fowler, 1991), and auditory streaming judgments (O'Leary & Rhodes, 1984). In most tasks, vision appears to exert a stronger pull; however, audition has been shown to be a strong force under certain conditions when making evaluations of tone duration (Walker & Scott, 1981). Cross-modal influences have been shown to be especially salient when the modality influenced (either audition or vision) is ambiguous (Wada, 2003).

### Cross-modal interactions (General)

Music-based studies have demonstrated "McGurk effect" parallels in music perception, albeit to a lesser degree than those found in speech. One such experiment showed that visual depictions of cello bowing and plucking motions can influence performance on musical tasks such as timbre identification (Saldaña & Rosenblum, 1993). Others have shown the visual mode superior to either the audio or audio-visual for conveying emotional expression (Davidson, 1994), and that visual information can influence expert musicians' judgments of certain aspects of performance quality (Gillespie, 1997).

### Cross-Modal Interactions (Percussion)

There is a long-standing debate within the percussion community concerning whether it is even possible to perform staccato and legato strokes on the marimba. A sequence of notes performed in a detached or disconnected manner would be considered staccato, whereas the same sequence of notes performed in a smooth or connected manner would be considered legato. Some percussionists believe that it is possible to control articulation on the marimba through changes in certain physical parameters of stroke (Bailey, 1963). Others remain adamant that these parameters in and of themselves play no direct role in altering the acoustical characteristics of individual notes (Stevens, 1990). Recent research using frequency spectrum analysis suggests there may be no quantifiable acoustical difference between tones produced by differing stroke types (Saoud, 2003). Perhaps the renowned percussionist Buster Bailey of the New York Philharmonic hinted at the possibility of cross-modal interactions in stroke type perception years ago when he remarked, "[Percussionists] should acquire a technique which will enable us to project a legato feeling when desired even though it is impossible to have a legato sound in the true sense of the word" (Bailey, 1963).

## 2. METHOD

In the process of creating stimuli for the present study, a world-class percussionist was asked to play a variety of tones with four different stroke types: normal, staccato, legato, and damped (i.e. the note was muffled shortly after being struck). One example of each stroke type was chosen for E1 (~ 82 Hz), A2 (~220 Hz), D4 (~587 Hz), and G5 (~1568 Hz). Although he was aware of the stimulus preparation procedure, he was only instructed to perform each stroke type as clearly as possible, without any additional instructions regarding expressive gestures. These samples were then used to generate the 76 'tokens' presented to subjects. In the context of this study, a 'token' is defined as a segment of audio, visual, or audio-visual information. The 76 tokens were divided into 3 categories: 48

audio-visual, 12 video-alone, 16 audio-alone (the damped stroke was only presented through the audio mode).[ii].

Subjects were current college students or recent graduates, grouped into three categories: non-music majors (24 subjects), music majors who did not declare percussion as their primary instrument (24 subjects), or music majors who declared percussion as their primary instrument (22 subjects).

Subjects responded to all questions using an unmarked 101 point slider. For all tokens, subjects were asked to respond by moving the continuous slider to indicate their perception of the stroke as more strongly 'staccato' or more strongly 'legato' (a value of 0 corresponded to a maximally staccato stroke, a value of 100 corresponded to a maximally legato stroke). In the audio-alone and video-alone modes, subjects were instructed to base their responses on the presented modality. In the audio-visual mode, subjects were clearly instructed to base their responses on the auditory information alone. In order to ensure they were attending to visual information while basing their answers on the audio, subjects were asked to make a second rating concerning the degree of mismatch between the audio and visual information. Previous experiments have demonstrated that asking subjects to provide such discrepancy ratings does not interfere with the primary task of judging the information itself (Rosenblum & Fowler, 1991). In order to maximize the study's musical validity, we did not manipulate the material other than uniformly normalizing the audio. Therefore all material used for this study was very similar to that which subjects might encounter when watching and/ or listening to a live performance.

# 3. RESULTS

A repeated measures ANOVA was conducted, using one between-groups variable (non-musicians, music majors, and percussion majors) and three within-subjects variables: pitch level (E1, A2, D4, G5), visual stroke type (legato, normal, staccato), and audio stroke type (legato, normal, staccato, damped). Because the data set violated the assumption of sphericity as calculated using Mauchly's test, we have adopted the more conservative Greenhouse-Geisser values in this report. All within-subjects factors proved to be statistically significant: pitch level ($F_{(2.42, 162..20)}$=196.18; $p$ < .0005), visual stroke type ($F_{(1.54, 103.40)}$=114.12; $p$ < .0005), and audio stroke type ($F_{(2.34, 156.66)}$=107.53; $p$ < .0005). In addition, there were two significant interactions: pitch level by group ($F_{(4.84, 162.20)}$=9.342; $p$ < .0005) and pitch level by auditory stroke type ($F_{(7.145, 478.71)}$=18.76; $p$ < .0005). Surprisingly, no significant interaction was observed between visual and audio stroke types: $F_{[6, 62]}$=1.842, $p$ < .109. The data do support our hypotheses that pitch level, visual stroke type, and audio stroke type exert a significant influence on subject ratings of staccato-legato. While it was not shown at a level of statistical significance, we did observe a trend that suggests visual information exerts an influence upon auditory perception of stroke type.

# 4. DISCUSSION

## Unimodal Examination

### Video-alone

When examining responses to the visual-alone tokens, a repeated measures ANOVA revealed no significant difference between groups (F(2,67)=1.269; p = .288). However, within-subjects variables revealed significant differences: pitch level (F(2.45, 161.83)=19.69; p < .0005) and stroke type (F(1.48,99.43)=257.05; p < .0005) were the statistically significant factors. Statistically significant interactions included: stroke type by group (F(2.97,99.43)=3.50; p = .019), pitch level by stroke type (F(3.99,267.40)=9.55; p < .0005), and pitch level by stroke type by group (F(7.98,267.40)=3.40; p = .001). When tested in the video-alone mode, subjects were rather adept at recognizing differences between stroke types. Legato, normal, and staccato notes were distinguished more clearly by vision, with more consistent ratings across different pitch levels than in the case of audio-alone (Figure 1).

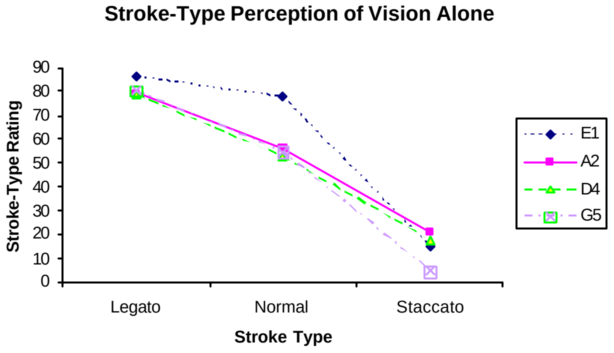

**Figure 1** Mean subject responses to the various marimba pitches by stroke type in the video-alone condition.

### Audio-alone

When examining responses to the audio-alone tokens, a repeated measures ANOVA revealed no significant difference between groups (F(2,67)=.184; p = .832). The same analysis revealed significant differences between all within-subjects main effects and interactions: pitch level (F(2.45,164.18)=121.56; p < .0005), stroke type (F(2.74,183.24)=92.59; p < .0005), stroke type by subject group (F(5.47,183.24)=2.99; p = .01), pitch level by subject group (F(4.90,164.18)=5.80; p < .0005), pitch level by stroke type (F(5.73,384.03)=14.02; p < .0005), and pitch level by stroke type by subject group (F(11.46,384.03)=2.40; p = .006).

For the three lowest notes (E1, A2, D4) legato, normal, and staccato strokes appear to be perceived as very similar in articulation when judged based on audio-alone, although damped strokes appear distinguishable from other stroke types. The highest note used in this study (G5) was rated far more 'staccato-like' than others, perhaps due to the characteristic lack of resonance in the highest register of the instrument. While legato, normal, and staccato strokes were difficult to distinguish within each pitch level, damped notes were rated consistently more 'staccato-like' for the three highest pitch levels (A2, D4, G5; see Figure 2).

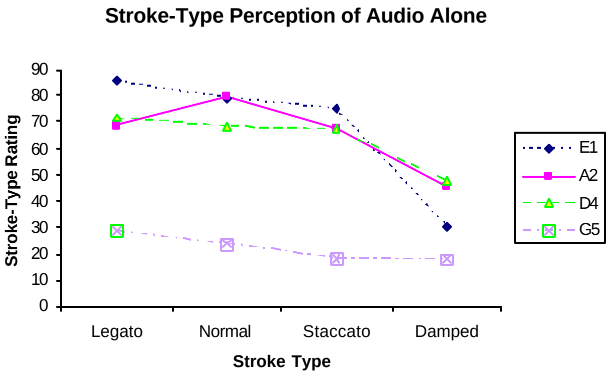

**Figure 2:**. Mean subject responses to the various marimba pitches by stroke type in the audio-alone condition

## 5. Bimodal Examination

Although subjects were asked to base their responses on the audio information alone, visual material appears to shift perception in a predictable manner when paired with identical audio samples (legato visual information causes a shift towards legato perception and vice-versa; see Figure 3).

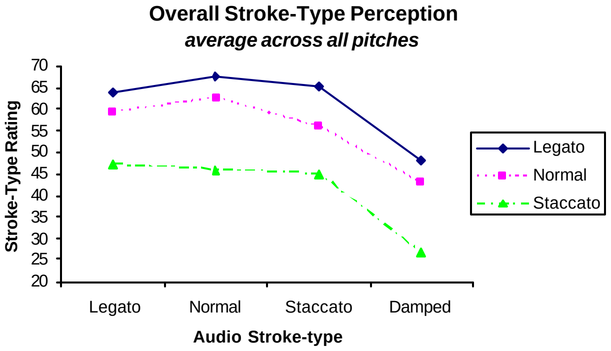

**Figure 3:** Depicts the effect of visual information on judgments of stroke type taken from audio-visual stimuli averaged across all pitch levels.

Though the influence of visual information on auditory perception was determined to be statistically non-significant at a .05 alpha level, there is a distinct trend observable in these responses that deserves the attention of further research. Comparison of mean ratings among different pitch levels reveals that articulation distinctions become more difficult to distinguish with increases in register. While in some cases there is confusion between legato and normal stroke types, the distinction between legato and staccato strokes is preserved across all four pitch levels. Within each individual pitch level, the distinction between legato, normal, and staccato notes was not as marked as the distinction between staccato and damped notes. Likely, this confirms our hypothesis that it is difficult to distinguish between the three undamped stroke types in the audio-alone mode. We continue to believe that in a real world live performance setting, visual information plays a significant role in auditory perception of stroke types as audio information alone is insufficient for making stroke type determinations. In the highest register, visual information may be the only method of distinguishing between stroke types since note length is so short as to leave little room for audible distinction. Recent studies have shown that visual information can alter the perception of auditory information when the auditory information is ambiguous (Wada, 2003).

By comparing the ratings of the audio information in an audio-alone token with the rating of the same information in an audio-visual token, it is possible to observe the influence of visual information on subject r atings. Figure 4 shows the differences in ratings of the same audio material when presented as part of an audio-visual token, versus as an audio-alone token. Examining the relative rankings of selected audio-visual examples can offer some insight into the mode exerting the strongest influence on auditory perception. Within each pitch level, tokens using a visual staccato stroke were perceived as 'more staccato' than tokens using a visual legato stroke, *independent of the audio with which it was paired.* As shown in Table 1, visual information consistently served as a better predictor of perceived stroke type for all four pitch levels.

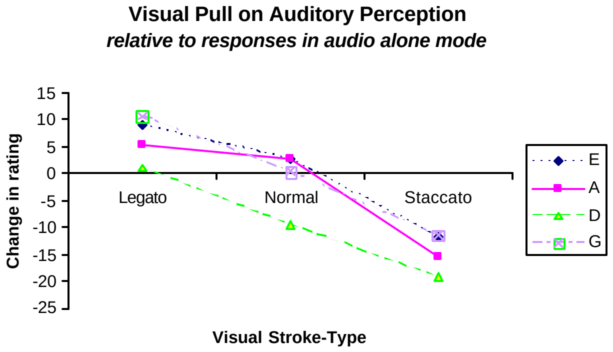

**Figure 4:** Visual information appears to exert a strong 'pull' on subject ratings despite repeated instructions for subjects to base responses solely on the auditory information in an audio-visual stimulus.

## 6. DISCUSSION

While our results did not reveal differences at a level of statistical significance, there is good reason to believe future research in this area will prove rewarding. Our investigations into cross-modal interactions of audio and visual information demonstrated a consistent effect of visual information on auditory perception, despite the fact subjects were briefed ahead of time as to the nature of the experiment and clearly instructed to base their responses on auditory information alone.

| *Pitch level* | *Shortest Rated Token* | | | *Longest Rated Token* |
|---|---|---|---|---|
| *E1* | Staccato<br>Staccato<br>(65) | Staccato<br>Legato<br>(68.6) | Legato<br>Staccato<br>(80.4) | Legato<br>Legato<br>(80.5) |
| *A2* | Staccato<br>Legato<br>(54.5) | Staccato<br>Staccato<br>(54.8) | Legato<br>Legato<br>(71) | Legato<br>Staccato<br>(77.3) |
| *D4* | Staccato<br>Staccato<br>(42.6) | Staccato<br>Legato<br>(47.8) | Legato<br>Legato<br>(68) | Legato<br>Staccato<br>(69.7) |
| *G5* | Staccato<br>Staccato<br>17.4 | Staccato<br>Legato<br>(17.8) | Legato<br>Staccato<br>(33.1) | Legato<br>Legato<br>(36.9) |

**Table 1:** Ordering of mean rankings for all audio-visual stimuli organized by pitch level. Visual stroke type is listed first and auditory stroke type second (e.g. the paring of a staccato visual and legato audio sample resulted in the audio-visual token rated most 'staccato-like' within the pitch level A2).

There are several potential reasons why our data did not prove to be of greater statistical significance. One case in which audio information has been shown stronger than visual is in making judgments of tone duration in the presence of discordant audio-visual information (Walker & Scott, 1981). While the present experiment addresses a slightly different question (that of stroke type as opposed to tone length), the two tasks share the key similarity of requiring sensitivity to temporal changes in auditory information. Therefore it is possible that similar experimental approaches to other instances of cross-modal interaction (measuring perception of dynamics, tempo, note density, timbre, etc) might yield more statistically conclusive results. In a sense, we were working in what could be considered a scientifically weak area of visual influence, yet were still able to demonstrate an observable trend of visual pull on auditory perception. Most importantly, we built a case through a variety of comparisons that in the presence of discordant audio and visual stroke type information, visual information is a stronger predictor of perceived stroke type, even when subjects were explicitly asked to base their ratings on the audio information alone.

Other music based studies have explored cross-modal interactions (Davidson, 1994; Gillespie, 1997), one even through a similar methodology (Saldaña & Rosenblum, 1993). Our approach was slightly different in that we did not obfuscate the audio material in any way. In order to keep our results as relevant as possible to practical issues of live musical performance, our demonstrations of the primacy of visual information are not based on cases where audio material has been altered in order to increase the level of ambiguity – nor was the visual material altered to heighten its clarity. Rather, we used audio recorded in a recital hall on a professional quality instrument performed by an internationally renowned marimba artist who was not coached ahead of time on his performance - he was simply asked to perform each of the four stroke types as clearly as possible.

Evaluations of stroke type are quite difficult to quantify scientifically. It is very rare musically to make judgments of stroke type based on isolated notes, and rarer still to listen to multiple notes and judge each independently. It is also unusual to methodically compare the stroke types of consecutive notes so drastically different in register. Each note on a marimba reacts differently, making consistent scientific evaluations of isolated notes (and therefore stroke type) a difficult task.

## 7. CONCLUSION

Much work remains to be done in exploring the role of visual information on auditory perception within the realm of live performance. Further studies of this nature would prove beneficial not only to the research community, but also to performers interested in incorporating musically useful gestures into their live performances. While much of our data did not prove statistically conclusive, we did succeed in demonstrating the fact that visual information can affect auditory perception of stroke type in a predictable manner. Similar research methods might be used in the future to investigate the role of visual information on the perception of other musical tasks such as loudness, timbre, tempo, or any number of other musical characteristics.

## 8. REFERENCES

1. Bailey, Buster. Mental and Manual Calisthenics for the Mallet Player. Warner Bros.: New York, pp v-vii, 1963.

2. Bertleson, Paul and Radeau, Monique. "Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations", Perception and Psychophysics 22(2), 1977.

3. Bertleson, Paul and Radeau, Monique. "Cross-modal bias and perceptual fusion with auditory–visual spatial discordance", Perception and Psychophysics 29(6), 1981.

4. Davidson, Jane. "Visual Perception of Performance Manner in the Movements of Solo Musicians", Psychology of Music vol 21, 1994.

5. Gillespie, Robert. "Ratings of Violin and Viola Vibrato Performance in Audio-Only and Audiovisual Presentations", Journal of Research in Music Education, vol 45 (2), 1997.

6. McGurk, H. & McDonald, J. W. Hearing lips and seeing voices. Nature, 264, 746-748, 1976.

7. O'Leary and Rhodes, Gillian. "Cross-modal effects on visual and auditory object perception", Perception and Psychophysics 35(6), 1984.

8. Rosenblum, L. D & Fowler, C. A. "Audiovisual investigation of the loudness-effort effect for speech and

nonspeech events", Journal of Experimental Psychology: Human Perception & Performance, 17, 976-985, 1991.

9. Saldana, Helena and Rosenblum, Lawrence. "Visual influences on auditory pluck and bow judgments", Perception and Psychophysics 54 (3), 1993.

10. Saoud, Erick. The Effect of Stroke Type on the Tone Production of the Marimba. Percussive Notes, 41 (3) 40-46, 2003.

11. Stevens, Leigh Howard. Method of Movement for Marimba. 2nd ed. Keyboard Percussion Publications: New Jersey, 1990.

12. Wada, Yuji, Kitagawa, Norimichi and Noguchi, Kaoru. "Audio-visual integration in temporal perception," International Journal of Psychophysiology, 2003 (in press)

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

14. Weerts, Theodore and Thurlow, Willard. "The effects of eye position and expectation on sound localization", Perception and Psychophysics 9(1A), 1971.

[i] The authors would like to acknowledge the collaborative assistance of Mr. Michael Burritt, a world-class percussionist and pedagogue at Northwestern University. It is due to his invaluable contribution to this project that the referenced high level of ecological validity was maintained.

[ii] Examples of the audio-visual tokens can be found on the ICMPC8 Proceedings CD-ROM under the following filenames: LegatoVisual_LegatoAudio.mpg, LegatoVisual_DampedAudio.mpg, and StaccatoVisual_StaccatoAudio.mpg. All stimuli are available online at: http://faculty-web.at.northwestern.edu/music/lipscomb/stimuli/.