CROSS-MODAL INTEGRATION IN MUSIC PERCEPTION

Michael Maksimowski and Frank Russo

Dept. of Psychology, Ryerson University, 350 Victoria St. Toronto, Ontario, Canada, M5B 2K3

1. INTRODUCTION

Music is an ideal domain through which to study issues concerning cross-modal integration. In addition to auditory information, the perception of music typically involves visual information and can in some instances involve vibrotactile information. Music is also a complex auditory stimulus; pitch, dynamics, mode and temporal information must all be integrated in order to generate meaning. According to the inverse effectiveness rule (Meridith & Stein, 1986), the greater the object ambiguity from a perceptual standpoint, the greater the opportunity for multiple modalities to enhance perception. Recent research has demonstrated a strong influence of visual information on auditory judgments concerning music (e.g., Thompson, Graham, & Russo, 2005; Thompson & Russo, 2007). Similarly, visual and vibrotactile information has been shown to influence speech recognition (Fowler & Dekle, 1991; Gick, Jóhannsdóttir, Gibraiel, & Mühlbauer, 2008). However, we have very little empirical information regarding integration of vibrotactile information in music. In the following experiment, participants made judgments of interval size for cross-modal presentations of intervals comprised of stimuli presented using audio alone, audio-visual, and audio-vibrotactile signals. In the latter two conditions, participants were instructed to base judgments on the auditory information alone. Results showed that accuracy of interval size was significantly greater in both the auditory-visual and auditory-vibrotactile conditions compared to audio-alone. Audio-visual and audio-vibrotactile conditions were not significantly different from one another. In light of these findings, differences in the extent of visual and vibrotactile influences on auditory judgments and the role of learning in cross-modal integration in music are discussed.

Music perception is fundamentally a multimodal experience. Yet the extent to which cross-modal integration in music is learned is relatively unknown. The primary purpose of this paper is to examine whether cross-modal integration in music requires learning. To test this, we compare the benefits of adding an additional modality of information to participants who are engaged in making judgments about the auditory stimulus: visual or vibrotactile.

Although visual features are frequently paired with auditory information in natural displays of music, the same is not true of vibrotactile information. Moreover, vibrotactile information tends to be dominated by low-frequency components. If we find that vibrotactile information supports music perception to the same extent as visual information, this would suggest that cross-modal integration in music is not entirely dependent on learning. We hypothesize that visual and vibrotactile signals, when added with audio, serve to enhance the information acquired.

2. METHOD

Ten university students (2 male, 8 female) participated in the experiment, ranging in age from 19 to 26 years. Years of formal musical training ranged from 0 to 7 years, with an average of 3.1 years (SD = 2.5). Only one participant reported current musical activity, and no participant reported any problems with hearing.

The music stimuli consisted of audio-visual recordings of 22 ascending intervals sung by two trained female vocalists. Each interval consisted of two tones in sequence. Each tone was approximately 1500 ms in duration and the pitch separation between tones ranged from 1 to 11 semitones. All tones fell within the range of 233.1 Hz (Bb3) and 440 Hz (A4). These intervals were presented in 1 of 3 ways: audio-alone, audio-visual, and audio-vibrotactile. Auditory information consisting of intervals and white noise (included to enhance difficulty) was played over headphones (Sennheiser HD580). Visual information was displayed on a 13-inch MacBook. Vibrotactile information consisted of vibrations applied to the palm of each hand using a pair of skin stimulators (Tactaid VBW32) with a useable output of 100 to 800 Hz and a peak frequency of 250 Hz. The perceived vibrotactile intensity of all stimuli was equalized based on a pilot experiment involving a separate group of participants. A channel box (M-Audio) was used to play both auditory and vibrotactile signals simultaneously.

Participants were seated at a desk in a well-lit double-walled sound-attenuated chamber (Industrial Acoustics Company). Trial presentation and responses were computerized using Experiment Creator X (Thompson & Kosalev, 2000). A within-subjects design was utilized; all participants began with the audio-alone condition and then proceeded to the audio-visual and audio-vibrotactile conditions. The order of presentation of the latter 2 conditions was counter-balanced across participants. At the beginning of each condition, participants were given practice intervals of 0 and 12

semitones. These intervals were not used in the actual experiment. Accuracy was calculated by taking the absolute difference between the participant's response and the actual semitone range (i.e., lower is more accurate). At the end of each trial, participants were prompted to press the key that corresponded with the semitone interval they just heard. All statistical analyses used an alpha-level of .05.

3. RESULTS

We evaluated the hypothesis that accuracy would improve in the bimodal conditions (audio-visual and audio-vibrotactile) relative to audio alone. Figure 1 displays mean accuracy (and standard error) for each condition while Figure 2 displays mean accuracy for each participant based on condition. A one-way within-subjects ANOVA showed that there was a significant effect of presentation condition, $F (2, 1317) = 12.701$, $p < .001$. Post-hoc analyses revealed that the significant differences existed between audio ($M = 2.58$, $SE = 0.102$) and audio-visual ($M = 2.05$, $SE = 0.083$) and audio and audio-vibrotactile ($M = 2.01$, $SE = 0.084$), but there was no significant difference between the audio-visual and audio-vibrotactile conditions.
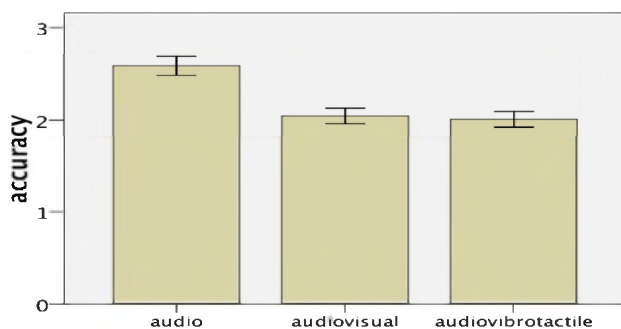

Fig 1. Mean accuracy levels for each condition, with error bars denoting standard error.
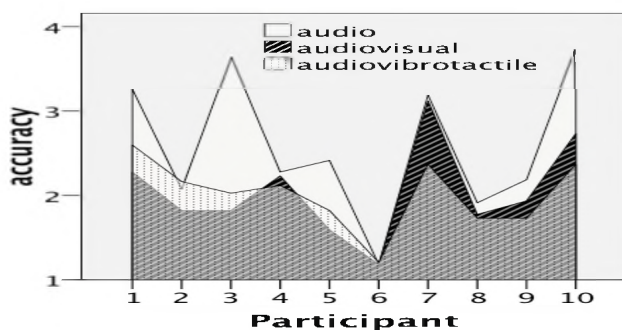

Fig 2. Participant-by-participant improvement in accuracy across conditions.

4. DISCUSSION

The results indicate that visual and vibrotactile inputs enhance accuracy for judgments of interval size compared to audio alone. Most importantly, however, is that participants gained equal enhancements of accuracy from both audio-visual and audio-vibrotactile conditions. Similar to speech (Fowler & Dekle, 1991), congruent vibrotactile information minimizes uncertainty and supports perception of the auditory signal.

Relative gains from vibrotactile exposure are not surprising considering that auditory-vibrotactile associations are both lawful and stem from same environmental signal (i.e., pressure waves). Cross-modal pairings of auditory and vibrotactile information are relatively unfamiliar to participants. Although vibration can be a desired aspect of music listening, the prominent vibrational components tend to be lower in frequency than those presented in the current experiment. Moreover, in ecological settings, lower- and higher-frequency vibrational components are available simultaneously, resulting in masking of higher-frequency components.

Our findings suggest that benefits of cross-modal integration in music are not *dependent* on learning. This view falls in line with direct theories of perception (Gibson, 1966; Liberman *et al.*, 1967). We do not suggest however, that integration of non-auditory information is devoid of learning in the case of music. Many visual influences in music are clearly best understood from a semiotic perspective (Thompson *et al.*, 2005). Future research should clarify mechanisms of integration and investigate the manner and extent to which vibrotactile information may support the musical experience.

REFERENCES

Fowler, C. A., Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 17* (3), 816-828.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Hougton Mifflin.

Gick, B., Jóhannsdóttir, K. M., Gibraiel, D., Mühlbauer, J. (2008). Tactile enhancement of auditory and visual speech perception in untrained perceivers. *Journal of the Acoustical Society of America Express Letters, 123* (4), 72-76. doi: 10.1121/1.2884349

Liberman, A. M., Cooper, F., Shankweiler, D., Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

Meredith, M. A., Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology, 56*, 640-662.

Thompson, W. F., Graham, P., Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica 156*, 177-201.

Thompson, W. F., Russo, F. A. (2007). Facing the music. *Psychological Science, 18* (9), 756-757.