Audiovisual integration in music performer recognition: do you need to see me to hear me?

Helen F Mitchell,[*1] Raymond AR MacDonald [#2]

[*] *Sydney Conservatorium of Music, University of Sydney, Australia*
[#] *Department of Psychology, Glasgow Caledonian University, UK*
[1]helen.mitchell@sydney.edu.au, [2]raymond.macdonald@gcu.ac.uk

ABSTRACT

Background

Listeners take for granted not only their capacity to distinguish between musical instruments, but also their ability to discriminate between performers playing the same instrument by their sound alone. Sound perception is usually considered a purely auditory process but seminal research in speaker recognition suggests that auditory and visual presentations of a person are integrated as each modality presents the same information, but in a different way. Listeners combine these cross-modal perceptions to recognise the person speaking (Kamachi, Hill, Lander, & Vatikiotis-Bateson, 2003; Lachs & Pisoni, 2004; Schweinberger, Robertson, & Kaufmann, 2007). In fact, more recent studies have suggested that listener/viewers are even more sensitive to the naturalistic timing of audio-visual speech presentation than first thought, and rely on seeing facial movements begin before audio commences (Maier, Di Luca, & Noppeney, 2011).

This has profound implications for music performer recognition, if multimodal information is combined to identify a particular performer.

Aims

The aim of this study is to investigate the integration of cross-modal sensory experiences (visual and auditory) in the recognition of individual music performers.

Method

Saxophonists (n=5) performed three jazz standards for an audio and video recording. Audio-only and video-only clips of the saxophonists' performances were prepared as stimuli. Listeners (n=20) attended a single cross-modal perceptual test and were assigned to a V-A or A-V group. In the V-A group, listeners watched a silent video clip of a saxophonist playing. Then, listeners heard two audio clips of the same piece, one by the performer they saw (target) and the other by a different saxophonist (distracter) and had to make a forced-choice decision about which audio clip was the original performer. For the A-V group, the test was reversed, and listeners heard an audio clip and were presented with two silent video clips. Finally, listeners were asked to rate their confidence in their responses and to explain how they made their choices.

Results

The majority of the participants were able to perform the cross-modal matching task above chance, regardless of the presentation order (V-A group M=68%, A-V group M=59%). Neither group was very confident in their choices (M=4.3/10). The number of correct responses by the V-A group was significantly greater then the A-V group ($p<.05$). Participants were more likely to recognise performers by ear after they had watched their performance.

Conclusions

This cross-modal matching task was designed to investigate if listeners can identify a performer's sound from the performer's visual cues. Participants were able to use the information about a performer in one modality and match it to the same performer in another modality. As in speaker studies, these listener/viewers were also sensitive to naturalistic timing and were more successful at hearing the performance after seeing it. These results will be discussed with reference to musical identities and sound recognition and will provide insights into the way auditory experts, such as musicians, identify individual musicians' sound.

Keywords

Music performance; musical identities; sound recognition, audiovisual integration

REFERENCES

Kamachi, M., Hill, H., Lander, K., & Vatikiotis-Bateson, E. (2003). `Putting the Face to the Voice': Matching Identity across Modality. *Current Biology, 13*(19), 1709-1714.

Lachs, L., & Pisoni, D. B. (2004). Crossmodal Source Identification in Speech Perception. *Ecological Psychology, 16*(3), 159 - 187.

Maier, J., Di Luca, M., & Noppeney, U. (2011). Audiovisual asynchrony detection in human speech. *J Exp Psychol Hum Percept Perform, 37*(1), 245-256.

Schweinberger, S. R., Robertson, D., & Kaufmann, J. M. (2007). Hearing facial identities. *The Quarterly Journal of Experimental Psychology, 60*(10), 1446-1456.