Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, July 23-28, 2012, Thessaloniki, Greece
Cambouropoulos E., Tsougras C., Mavromatis P., Pastiadis K. (Editors)

The Coupling of Gesture and Sound: Vocalizing to Match Flicks, Punches, Floats and Glides of Conducting Gestures

Aysu Erdemir,[1] Emelyne Bingham,[2] Sara Beck,[1] John Rieser[1]

[1]*Psychology and Human Development in Peabody College, Vanderbilt University, USA*
[2]*Blair School of Music, Vanderbilt University, USA*

aysu.erdemir@vanderbilt.edu, emelyne.m.bingham@vanderbilt.edu, saralynbeck@gmail.com, j.rieser@vanderbilt.edu

ABSTRACT

Background

Past studies have shown the importance of physical movement for musical communication by showing that audiences can perceive expressive (Davidson, 1993), emotional (Dittrich et al., 1996), and structural characteristics (Schutz & Lipscomb, 2007) of music through vision alone. This study adds to the literature by exploring cross-modal interactions of movement and sound across persons.

Aims

The study was designed to explore whether there was a systematic relationship between various hand gestures performed by an expert conductor, and accompanying vocal sounds produced by adults with or without any kind of musical background. We wanted to see whether people automatically and systematically vary their utterances in a way to match the movement characteristics of certain gestures. For this reason, we picked gestures that are not contained in conducting manuals and in most conductors' repertoire, but nevertheless seem familiar/natural in an everyday life context. As a follow up, we quantified the psychoacoustic properties of the sound data, so that we could relate each of the four gestures to specific acoustic features that correlated with them.

Method

Participants (n=36) some with and others without musical training, watched videos of four different right hand gestures called "flicks, punches, floats and glides". Each video consisted of a conductor performing each gesture 10 times in a row at a constant tempo. The gestures were part of Effort Actions proposed by Laban (1996), and varied in terms of their use of space (direct/indirect), weight (strong/light) and time (sudden/sustained). In an effort to isolate the hand gestures and remove emotional cues from the face, the videos contained only the conductor's chest, shoulders, arms and hands in view. Participants were asked to produce the syllable /dah/ repeatedly in a way that feels natural to the four gestures they observed visually. No other instruction was given. The task was to speak /dah/ instead of singing a melody because some of our participants lacked musical background; and acoustical analysis could be performed more reliably on a single sound than sung melodies. In order to test our hypothesis, three independent judges listened to and scored the audio-recordings of the vocal responses without any knowledge of the visual gestures, and judged which of the four gestures gave rise to the produced dahs. We defined accuracy as correct categorization of the sounds into one of four movement categories.

In order to further explore the link between the gestures and sounds, we performed a psychoacoustic analysis on the sound data, by extracting the acoustic features including duration (ms), amplitude (dB), fundamental frequency (Hz), pitch variability (Hz), and formant frequencies (F1, F2, and F3), which are then related to motion characteristics of the gestures.

Results

Results showed that categorization accuracies were 94%, 96%, 80%, and 82% for flicks, punches, floats, and glides, respectively. The categorization accuracies were above 84% on average even for those subjects with no musical background (n=12). Binomial tests revealed that percentages of correct classifications were significantly greater than the chance level of %25 for all four gestures (p<.00). Judges' most errors involved substituting flicks & punches, and substituting floats & glides, which fits kinematic similarities among gestures.

Additional audio analyses revealed significant associations of the motion characteristics of the gestures such as time, weight and space to overall duration, loudness & pitch levels of the utterances. Specifically, sudden vs. sustained gestures aroused shorter vs. longer responses; light vs. strong gestures aroused softer vs. louder responses; and direct vs. indirect gestures aroused sounds with lower vs. higher fundamental frequencies. Moreover, we observed systematic modifications in vowel formant frequencies for sudden vs. sustained gestures. In sum, participants varied their utterances systematically for each gesture being watched; indicating that observed motion characteristics modulated vocal responses in a systematic way.

Conclusions

The data collected imply a definable cross-modal relationship between gesture and sound, where the visual effects from the kinematics of movement patterns are automatically translated into predictable auditory responses. This reflects a strong coupling between motor and perceptual processes, where the motor system influences the perceptual interpretation of dynamic visual stimuli.

Keywords

Conducting, Gestures, Psychoacoustics, Music and Movement, Cross-Modal

REFERENCES

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psych of Music*, 21,103–113.

Dittrich, W. H., Troscianko, T., Lea, S. E. G., & Morgan, D. (1996). Perception of emotion from dynamic light-point displays represented in dance. *Perception*, 25, 727-738.

Laban, R. (1966). Annotated and edited by Lisa Ullmann. Choreutics. London: Macdonald & Evans.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, 36, 888-897.