



Super-Normal Integration of Sound and Vision in Performance

Joel D.F. Lim, Suzy J. Styles

Division of Psychology, School of Humanities and Social Sciences,
Nanyang Technological University, Singapore

joel0008@e.ntu.edu.sg, suzy.styles@ntu.edu.sg

Abstract

In the world of rock, blues, and popular music, guitarists who pull faces when they play are somehow more entertaining to watch – as if there is an agreement between the performer and the audience that certain facial gestures enhance the expression of the music at key moments. How does this agreement come about? What is it that viewers and performers share? Research in the field of cross-modal perception has demonstrated that humans share systematic linkages between the senses, including vision and audition. For example, people link higher pitched sounds with smaller, paler, and higher-up objects than lower sounds (e.g., Mondloch & Maurer, 2004). Yet few investigations have explored how these linkages relate to performance in the creative arts. One theory suggests that artists have stronger functional connectivity between sensory areas, and are therefore able to invoke or translate sensory experiences across modalities, using sensory metaphors, and strategic sensory exaggeration (Ramachandran, 2003). To test if viewers share a sensory experience with performers, we asked participants in an online quiz to guess which of two guitar faces ‘went with’ particular moments in guitar solos. One-hundred and seventeen volunteers guessed guitar faces at rates higher than would be predicted by chance, indicating that people were able to draw meaningful information about music from still images of the guitarist’s face. In addition to presenting the results of the study in a poster format, this demonstration will give visitors a chance to engage with the interaction between musical performance and performative gesture, in two activities.

Keywords: musical performance, orofacial gesture, crossmodal correspondence.

1. Introduction

Some of the best guitarists worldwide contort their faces unconsciously as they perform their solos, stretching and twisting their faces, sometimes to amusing effect. Compared to a deadpan performance, these contortions seem to make for a more entertaining audience experience. We know this phenomenon as ‘guitar-face’.

Music and gesture

Humans appear to be naturally inclined to move their bodies to music. Infants sway, move and dance to rhythms (Zentner & Eerola, 2010). Even ‘sophisticated’ adult concertgoers respond to the dynamics of musical

performance by dancing, nodding their heads, or making gestures along with the music, with face or limbs. These music-related gestures demonstrate the intrinsic sensory linkages between audition and the somatosensory experience of motion.

While we often think of performers as trained specialists in their chosen field, in “air guitar championships”, contestants compete to best co-ordinate sound-based gestures and physical expressions at a complex level. It is note-worthy that some of the most successful “air guitars” have not played actual guitars before (Godøy, Haga & Jensenius, 2006). As

such, there appears to be a shared agreement of what a good guitarist looks like when they play certain sounds, regardless of musical background.

The spontaneous production of sound-based gestures point to certain similarities in the way musical sounds are understood by both the novice and the expert: since audiences and air-instrumentalists are able to appreciate and/or emulate sound-based performance gestures, this suggests a sensory link between the auditory and the visual components of a live musical performance.

When both the performer and the audience share a sensory mapping between the two, then their shared experience may affirm, or even enhance the dynamics of musical performance, as it is being played.

Cross-modal correspondences

In the field of cross modal research, natural mappings between sensory experiences have been well documented. For example, both adults and 5-year-old children have been shown to match the intensities of loudness and brightness across modalities (Bond & Stevens, 1969), and adults and infants match higher pitched tones to smaller, paler, objects in higher spatial position (Mondloch & Maurer, 2004). In the domain of linguistic sound symbolism, Edward Sapir observed that majority of people tend to associate nonsense words like “mal” with large objects, and “mil” with small objects; pointing out an association between the vowels /a/ and /i/ and the size of an object (Sapir, 1929).

Ramachandran and Hubbard (2001) proposed that this kind of sound symbolism arises out of cross-modal correspondences between not only sensory maps (e.g. vision, audition) but also the motor maps used for speech. Using the nonsense words “Bouba” and “Kiki”, they found that majority of people pair the word “bouba” with rounded shapes, while pairing the word “kiki” with sharp, angular shapes. The authors suggested that the rounded shape of the lips in the articulation of a word-form like ‘bouba’ gives rise to the association with curved forms. Hence, natural

mappings may exist between facial gestures and features of the audio-visual landscape – in particular, the round lips of ‘oo’ go with curves, and with low sounds, while the thinly spread lips of ‘ee’ go with spikes and high notes.

Synaesthesia and the artist

Ramachandran and Hubbard (2001) proposed an explanatory model of cross-modal correspondences in language, using synaesthesia as a lens. Synaesthesia is a condition in which the stimulation on one sensory modality results in more than one sensory experience – for example, coloured hearing or tasting shapes (Harrison & Baron-Cohen, 1997), Ramachandran and Hubbard proposed that synaesthesia arises out of exaggerated patterns of cross-talk and cross activation between sensory brain regions (Ramachandran & Hubbard, 2001), and that this extra sensory excitation may make an individual more likely to be adept at connecting seemingly unrelated concepts. In his BBC Reith Lecture: “Purple Numbers and Sharp Cheese”, Ramachandran pointed out that synaesthesia is common among artists, writers and poets, whose works display skillful pairing of concepts, emotions, or sensory experiences that appear to be unrelated. It is suggested that while the neurological knack for understanding metaphors is present in all of us, the creative side is much stronger in artists and synaesthetes due to stronger cross-activation.

The creative arts utilize the vehicles of exaggeration, and distortion of stereotypes to trigger sensations of familiarity and surprise in the audience. For example in caricature drawings, unique features of a person’s face are emphasized and shifted away from the facial norm, until these exaggerations seem more like the person than the actual person themselves (Ramachandran, 2003). Similarly, in animal behavior experiments, Tinbergen and Perdeck (1950) demonstrated that creating an exaggerated combination of visual features could create a “supernormal” stimulus that would elicit greater feeding responses in gull chicks, than would their own parents’ beak. It may follow then that artists and performers who successfully entertain their viewers, have

managed to create super-normal sensory excitation: performances that successfully combine sensory stimulation across different modalities may appeal to the hardwired nature of cross-modal human perception.

Despite the persuasive nature of these ideas, the notion of supernormal inter-sensory stimulus in artistic communication has not yet been translated into a testable hypothesis (c.f., Styles, 2016, this volume).

Here, we ask the very testable question: If guitar soloists incorporate guitar-faces (visual) into their instrumental performances (auditory), and if audiences implicitly integrate the combination of information from different modalities, *and if those senses combine in a satisfactory manner* (i.e., if the guitar face is good), then it follows that the audiences will be able to uncover something about the music from the face that was pulled at the time.

To test this hypothesis, we conducted an online quiz to test whether people experience cross-modal correspondences for arbitrary selected moments from live guitar solo performances, by seeing if they are able to match the correct guitar-faces to their respective moments in the guitar solo, at rates above chance.

2. Experiment

We analysed the performance of 117 volunteers who underwent an online guitar-face quiz hosted on a Brain, Language and Inter-sensory Perception (BLIP) Lab webpage (URL <http://blogs.ntu.edu.sg/blip/>). Participants ranged from 18-67 years of age and included both people with (49.6 %) and without (50.4 %) prior music experience or training in music theory or performance. Of these, 46.2 % have experienced playing a guitar. Participants resided in various countries.

Figure 1 shows a sample layout for a single quiz item. Participants were presented with a guitar solo, and were asked to pay attention to a specific moment marked by the sudden appearance of a blue box on a time progress display. A short 100ms sound slice from the moment of the guitar-face was also provided for them to play for their reference. Two faces were displayed side-by-side (one target and one distractor) and participants were instructed to guess which of the faces was produced at the specified moment of the guitar solo, for 28 quiz items.

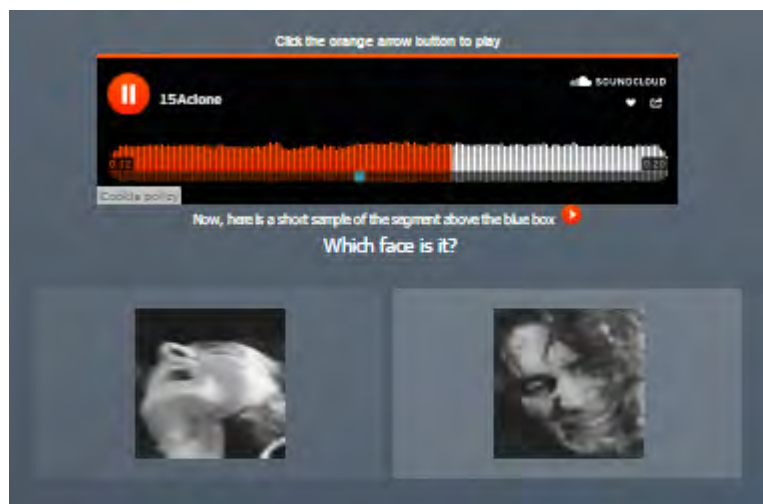


Figure 1. Example quiz question with scrolling music bar at top, guitar-face 'moment' marked by a blue box in the timeline, and two faces drawn from the same solo.

3. Results

Our results show that participants are able to draw meaningful information about the musical performance by simply looking at the guitar-face and therefore be able to correctly match it to its corresponding moment on the contour of the musical solo above chance levels ($M = 15.12$, $SD = 2.95$, $t(116) = 4.108$, $p < .001$).



Figure 2. The most recognisable guitar face (left) with its distractor on the right.

Notably, the most recognisable guitar-face (see Figure 2), had a facial gesture resembling the articulation of the vowel sound 'oo' (/u/). The corresponding moment for the target guitar face was preceded by a downward undulation of the music contour, and followed by an upward inflection of the pitch contour. This "valley" of low pitch in the music contour constitutes a compelling cross-modal correspondence to lip shape, and may form a type of super-stimulus.

Discussion

Our results show that in spite of being presented with no information about other contextual and visual aspects of the performance, the ability of the participants to match the guitar-face correctly with their corresponding moments still remained above the chance level.

The fact that participants were able to draw meaningful information and make pinpoint deductions at about the musical performance simply by looking at a guitar face and tracing the contour of a musical solo, suggests a strong correspondence between the motor maps underlying facial expression and the dy-

namics of music in live performance, which may then provide the building blocks for a supernormal stimulus in live performance; created and conveyed by the performer to be received and appraised by the audience. This may explain a shared experience between performer and audience that is multi-modal in nature; the performer and the artist have a natural basis for communicating and understanding linkages between facial gestures and musical dynamics.

5. Si15 Demonstration

In August 2015, we gave a public demonstration of the research at Singapore's ArtScience Museum, organised by Si15 Soundislands Festival (URL <http://soundislands.com/si15>). A gallery space was designed to engage members of the public in two interactive explorations of co-music gestures: An ongoing interactive data array (the guitar-face board), and a live air-guitar competition.

Collective guitar-face data array

The demonstration included the opportunity to listen to a series of guitar solos, accompanied by time-progress bars, with selected 'moments' of performance for evaluation. People were asked to select from a fixed set of schematic face shapes, which face 'goes best' with the performance for evaluation. People were asked to select from a fixed set of five schematic face shapes, and to indicate their response using one of five rubber stamps, on a response slip. These response slips were compiled into a live display.

Each face-shape stamp was assigned a different ink colour, such that the growing display of public selections created a colourful data-array as the day progressed. The interactive nature of the exhibit provided a tangible insight into how much an individual's sensory experience is shared with other members of the general public.

Guitar-Face Competition

Towards the end of the demonstration, a competition was held, with performers asked

to focus particularly on creating the best *guitar-face*, as highlighted by performing with their face inside a picture frame. This competition provided a light-hearted collective experience of what makes a 'good' guitar-face.

Acknowledgements

This project was supported by a Nanyang Assistant Professorship Grant (4081215) to SJS: *The Shape of Sounds in Singapore*.

6. References

- Bond, B., & Stevens, S. S. (1969). Cross-modality matching of brightness to loudness by 5-year-olds. *Perception & Psychophysics*, 6, 337–339.
- Godøy, R.I., Haga, E., & Jensenius A.R. (2006). Playing "air instruments": Mimicry of sound-producing gestures by novices and experts. In S. Gibet, N. Courty, & J.-F. Kamp (Eds.), *Gesture in Human-Computer Interaction and Simulation: 6th International Gesture Workshop, GW 2005, Berder Island, France, May 18-20, 2005, Revised Selected Papers, Volume 3881/2006*, pp. 256–267. Berlin Heidelberg: Springer-Verlag.
- Harrison, J.E., & Baron-Cohen, S. (1997), *Synaesthesia: A review of psychological theories*, in Baron-Cohen, S & Harrison, J.E., (Eds.), *Synaesthesia: Classic and Contemporary Readings*, pp.173-182. Oxford, England: Blackwell.
- Mondloch, C., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective & Behavioral Neuroscience*, 4(2), 133-136.
- Ramachandran, V.S. (2003). Purple Numbers and Sharp Cheese. Retrieved April 10, 2015, from http://downloads.bbc.co.uk/rmhttp/radio4/transcripts/20030423_reith.pdf
- Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia—A window into perception, thought and language. *Journal of Consciousness Studies*, 8, 3–34.
- Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, 12, 225–239.
- Styles, S.J., (2016), The language of dance: Testing a model of cross-modal communication in the performing arts. *IMCA Array, Volume 2016, Special Issue: Proceedings of Si15*.
- Tinbergen, N. & Perdeck, A. C. (1950). On the stimulus situation releasing the begging response in the newly hatched herring gull chick (*Larus argentatus* Pont.). *Behaviour*, 3, 1–39.
- Zentner, M. & Eerola, T. (2010). *Proceedings of the National Academy of Sciences*, 107(13): 5768–5773.