array

The Language of Dance: Testing a Model of Cross-Modal Communication in the Performing Arts

Suzy J. Styles

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

suzy.styles@ntu.edu.sg

Abstract

Integration between the senses is an intrinsic part of the human condition. Many forms of artistic expression make use of these sensory alliances, for example, the expression of rhythm and melody in dance. To test whether performers can effectively communicate information from one sensory modality (hearing) into another (vision), we asked one experienced dancer to perform dance-motions to the sounds of meaningless speech, and asked junior dancers to guess which dance motions were produced in response to which sounds. The junior dancers were substantially better than chance in the guessing task, suggesting that the dance performer successfully captured acoustic information about the identity of the speech sounds in her motions. We also found that dance experience did not predict performance in the task, suggesting that sensory congruence may not be learned through practice, but may be shared among the general population. However, a subset of dancers were much better than the main group, suggesting that sensory congruence may be differentially distributed through the population. This fits well with a model in which the strength of sensory connectivity differs across the population, and in which the creative arts attracts those individuals for whom the intrinsic links between the senses are experienced more powerfully.

Keywords: cross-modal perception, performing arts, sound symbolism

1. Introduction

Information floods our senses simultaneously, giving us an intensely multi-modal experience of our world. These sensory co-activations allow us knowledge about how animals with louder, lower-pitched calls have larger bodies (Morton, 1977), and that lowering certain resonant vocal frequencies generates dominance gestures (e.g., Chimps, 'ooh ooh' dominance versus 'ee ee' submission calls: Ohala, 1994).

The implicit understanding that certain kinds of sensory information 'go with' others is the tip of an evolutionary iceberg known as 'cross-modal correspondences'. For many years, researchers in the psychological sciences have experimentally demonstrated that high pitches are typically mapped to smaller, paler, spikier shapes, positioned higher up in visual space (Evans & Treisman, 2010; Spence, 2011), effects that are shown by chimpanzees (Ludwiga, Adachid, & Matsuzawa, 2011), as well as by human infants (Mondloch & Maurer, 2004; Walker et al., 2010). These findings demonstrate that connections between the senses are evident from the beginnings of human development, and perhaps most importantly for the creative arts, these correspondences are shared.

This means that sensory relationships for one person should more-or-less align with sensory relationships for the community at large: Our sensory system provides a platform for communication *across the senses*.

One off-shoot of cross-modal processing arises in the domain of language. Despite the ability of language to use arbitrary sounds to encode abstract meanings (de Saussure, 1911:1959), there is a growing body of evidence showing that certain speech sounds 'go with' certain sensations in other modalities. For example, when given two novel names for two novel objects, (one curvy, one spiky), the majority of people have a strong preference to match up the curvy shape with 'round-sounding' word-forms like 'bouba' or 'maluma', and spiky shapes with 'hard-sounding' word-forms like 'kiki' or 'takete' (Köhler, 1929:1947; Ramachandran & Hubbard, 2001).

This bouba/kiki task has been replicated in a variety of languages, for participants at various ages, including infancy (e.g., Bremner et al., 2013; Ozturk, Krehm, & Vouloumanos, 2013; Spector & Maurer, 2013). As such, sensory congruence for speech sounds forms part of the constellation of shared inter-sensory mappings.

Where sensory integration differs dramatically from the general population, we tend to think of this as dis-orderd. For example, people who experience 'synaesthesia' may spontaneously see colours and shapes when they listen to music, or experience colours when they see particular letters in printed text. The particular sensory experiences differ between individuals, but within an individual the sensations remain stable throughout adulthood (for overview, see Cytowic & Eagleman, 2009). Interestingly, synaesthetes tend to be over represented in the arts (Rothen & Meier, 2010).

Ramachandran and Hubbard (2001) proposed that synaesthesia might exist at the extreme end of normal sensory 'connectivity' or cross-modal integration, suggesting that inter-sensory connectivity is so strong it generates a kind of automatic sensory 'cross-talk'. This model of sensory congruence suggests that although we are biologically endowed to experience sensory connectivity, the extent of functional connectivity differs across the population; with dis-regulated hyper-connectivity at one end of the spectrum (i.e., synaesthesia), and dis-regulated hypo-connectivity at the other (e.g., sensory integration deficits).

Recent evidence from brain imaging studies has indeed suggested that people who show greater sensitivity to linguistic sound symbolism also exhibit greater activation in left superior parietal cortex, a brain region associated with cross-modal integration, as well as hallmarks of greater functional connectivity in the left superior longitudinal fasciculus (Pirog Revill, Namy, DeFife, & Nygaard, 2014). This finding lends support to the idea that differences in brain structure and function underpin differences in the strength of sensory integration across a population.

At the extreme end of the integration spectrum, several studies have also shown differences in brain structure and function for synaesthetes, as compared to the normal population (for overview, see O'Hanlon, Newell, & Mitchell, 2013). At the other end of the spectrum, people with autism show lower sensitivity to the bouba/kiki effect (Oberman & Ramachandran, 2008; Occelli, Esposito,
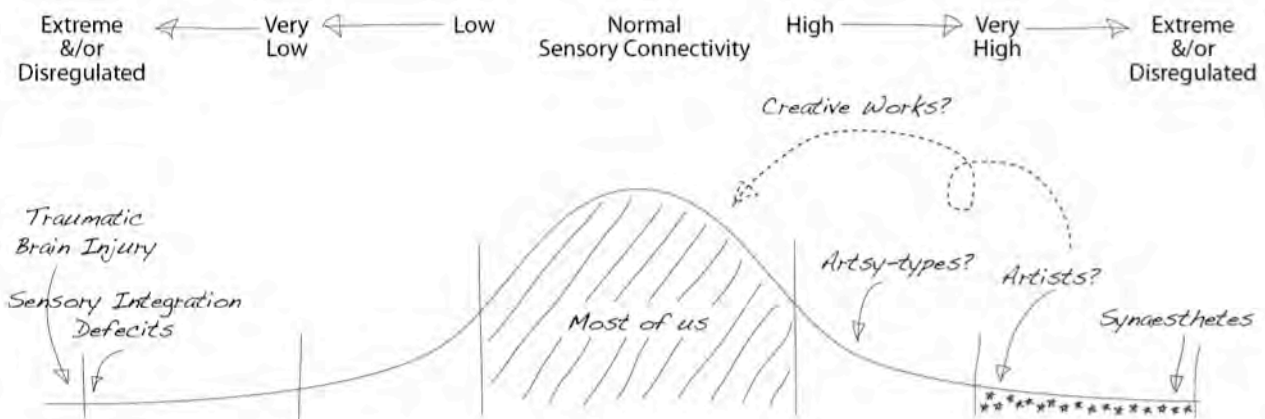
Figure 1. A population model of sensory integration strength. Normal distribution curve showing 1, 2, and 3 standard distributions from mean.

Venuti, Arduino, & Zampini, 2013), and this may be a marker of sensory integration deficits which are frequently reported as co-occurring with autistic traits. Taking these sources of information together, the idea of individual differences in sensory integration across the population seems well supported by the current evidence. A model of population differences in sensory connectivity is sketched in **Figure 1** (previous page).

This leads us to draft a model of sensory integration strength and creative arts across the population. In his public lectures (2003), Ramachandran has suggested that artists may be those in a community with greater sensory connectivity, which provides them with an enhanced ability to generate novel sensory metaphors (e.g., a sweet melody, a dark expression, angry brush strokes), but since their sensory congruence is part of the system shared by the population, these novel metaphors can be readily interpreted by the majority.

If we follow these suggestions to their logical conclusion, we can formulate a model of how certain creative arts relate to sensory integration strength, which gives rise to some interesting predictions: Firstly, there should be more consumers of art than creators, as the majority of us are endowed with sufficient sensory congruence to appreciate a well-made sensory alliance, but perhaps not the depth of sensory integration to imagine a novel relationship we have not yet encountered.

Secondly, there should be a sector of the community who are more deeply inter-sensory than most – they may be 'artistically inclined' but perhaps not as deeply engaged as artists (see **Figure 1,** 'Artsy types'). These people would be predicted to get more enjoyment than normal out of well-matched cross-modal experiences, and may actively seek them out: They may visit galleries, sing in choirs, dance in an ensemble, or perform in an orchestra, but they might be less inclined to create, conduct, choreograph, compose, or improvise.

Thirdly, there should be a corresponding sector of the community for whom sensory integration is lower, and may simply be less compelling. People in this group may be 'artistically disinclined', and have little interest in the non-literal components of art (e.g., "I like pictures where you can tell what it's supposed to be", "My three year old can drawn better than that").

Finally, we should predict that when an artistic expression 'works,' it is because the creative artist is able to communicate something (a percept, an emotion) through their chosen medium, in such a way that an audience is able to 'get it'.

The current study

Here, we used the medium of dance to see whether an artist (an experience dancer) could translate sensory information from one modality (sound) into another (sight), by generating novel dance movements to abstract sounds of speech. If our dancer can create motions which tap into sensory mappings shared with the general community, then other viewers of her motions should be able to extract relevant information about the sounds she had heard. If, on the other hand, her motions are too idiosyncratic (e.g., too synaesthetic), or don't contain sufficient sensory information (e.g., too weakly integrated), they might not contain sufficient shared sensory information for her audience to identify the sounds. In this small-scale investigation, we worked with junior dancers at a local high school.

2. Method

To make our speech sounds highly contrastive, we picked two of the most commonly reported syllables from the sound symbolism literature: /bu/ and /ki/ (e.g., D'Onofrio, 2014; Ramachandran & Hubbard, 2001). To give the dancer a more dynamic performance range, and to give viewers two types of acoustic information to base their judgements on, each syllable was produced in three rhythmic styles: a *short*, single articulation, a *long*, single articulation, and a *staggered* string of five short articulations, in which each short articulation was the same length as the short condition, and the total string was the same length as the long articulation (**Table 1**).

Table 1. Speech stimuli: two syllable types produced in three rhythmic patterns.

	Rhythm		
Speech	short	long	staggered
syllable	bu	bu~~~~	bu bu bu bu bu
	ki	ki~~~~	ki ki ki ki ki

A Singaporean bilingual speaker of English and Chinese produced each speech sound in each rhythm condition three times, using an animated gif as a timer. The clearest token of each sound type was selected for use in the test.

Speech-to-dance: motion generation

An experienced adult dancer was invited for a video recording session. She had several years of experience in Hip Hop and Street dance, with a current specialisation in Pop-n-lock. A single recording session was conducted at Nanyang Technological University.

The dancer was asked to produce one novel dance gesture for each of the six recorded sounds. The purpose of the experiment was explained, so that the dancer understood that later viewers would be guessing which movement was triggered by which sound. She was also instructed to use only meaningless, abstract motions based on the sounds she heard, rather than trying to depict meaningful shapes (e.g., tracing a letter shape). The dancer wore plain clothes appropriate for street dance. To remove the possibility that the dancer might inadvertently signal the identity of a speech sound with her face, the dancer wore a plain white mask.

All the speech sounds were played before video recording began, so that the dancer could familiarise herself with the sounds. One dance motion was recorded for each of the sounds. The dancer began each motion in the same position, and returned to that position at the end of her motion. She was allowed to review each recording, and re-take any video as many times as she liked. When the dancer was satisfied her motions expressed the sounds she had heard, the videos were trimmed into clips (see **Figure 2**), and edited to remove audio.
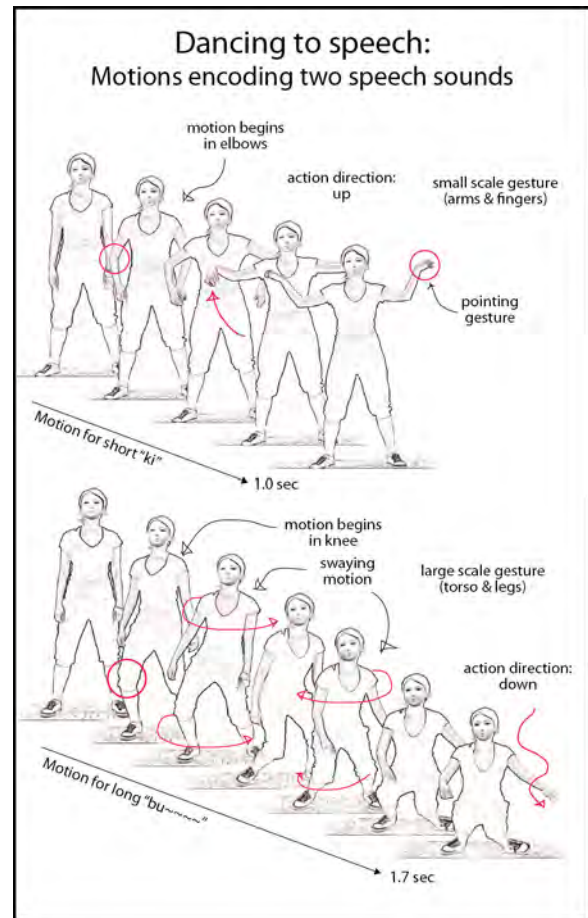
Figure 2. Illustration of two dance gestures produced in response to speech. Gesture durations measured from first frame of motion to last frame before returning to starting position.

Dance-to-speech: cross-modal matching

Participants in the cross-modal matching study were tested in three group sessions. Each dance motion was projected in front of the group, in silence, in a random order. Participants guessed which dance motion was produced in response to which speech condition, and marked their answers on a paper form. Before beginning the projection, participants had a chance to hear what each of the speech tokens sounded like. At the end of the experimental session, participants filled out a survey about their dance backgrounds. The whole procedure lasted a few minutes.

Junior dancers involved in dance classes at Nanyang Girls' School took part in the study after obtaining parental consent. The 20 par-

ticipants were aged 13-18 years, and had dance experience ranging from 5 to 14 years. There were no reported hearing problems, and normal (or corrected-to-normal) vision. The study was conducted at the regular meetings of three dance classes held at the school. The junior dancers had no contact with our experienced adult dancer. The procedure was approved by the NTU Institutional Review Board, in collaboration with Nanyang Girls' School.

Predictions

We predicted that if our dancer was able to capture something about the sounds of speech in her motions *and if other students of dance share a similar understanding of the links between sounds and motions*, then junior dancers should be better-than-chance at guessing which motion was produced in response to the different speech sounds, even if their dance backgrounds were different, and even though abstract, meaningless speech is not normally used as a stimulus for dance.

Since the rhythm classes differed substantially in their large-scale temporal structure, and rhythm is typically incorporated into dance performance, we predicted that participants might be better at matching dance motions to rhythm categories, than they would be at matching dance motions to speech sounds (which differ in their fine spectral and temporal detail).

We also predicted that if acoustic properties of the different speech sounds shared a sensory congruence with the rhythm classes they were produced in, that some of the motions would be more guessable than others, because they represented a kind of superstimulus, in which each feature of the articulation enhanced its identity, making it easier to generate a holistic motion for the sound.

Finally, since dance experience generates functional changes in the way dancers' brains respond to the motions of other dancers (Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005), we predicted that experience might modulate gesture/sound congruence, giving more experienced dancers better performance in the guessing task. On the other hand, even though dancers use gesture/sound congruence when they perform, it is possible that they generate motions which have a shared visual congruence for everyone, and practical experience in dancing may have little influence on how well dance viewers are able to match motions to sounds.

3. Results

Since each dance motion could be matched to any of the six speech stimuli, participants had a one-in-six chance (18%) of guessing correctly if they were picking purely at random – effectively the same as rolling dice. We found participants were significantly better than chance at picking the correct dance motion, with an overall accuracy of 56% ($t(19)=5.8$, $p<.001$, $d=1.21$ **Figure 3, Left**)
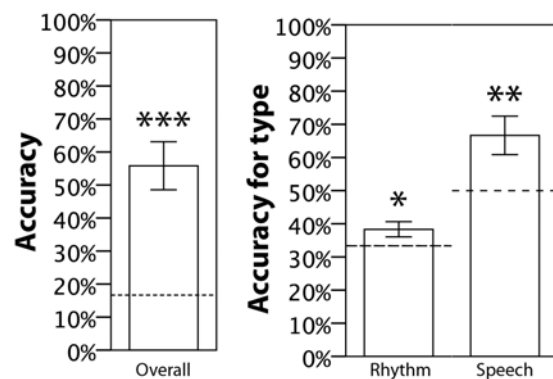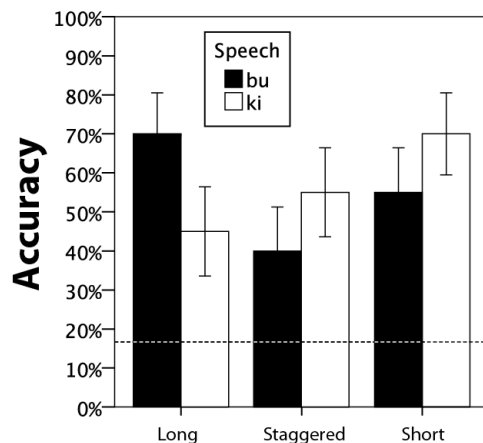

Figure 3. Average number of correct guesses per participant (**Left**), and shown separately for different acoustic features (**Right**). Chance values are dashed lines. Error bars +/- 1SE, ***p<.001, **p<.01, *p<.05.

We also looked separately at accuracy for rhythm, and accuracy for speech sound (**Figure 3, Right**). When allocating a dance motion to a rhythm category (regardless of which speech sound), participants were correct 38.0% of time, and this was significantly better than the chance value of one-in-three ($t(19) = 2.18$, $p < 0.05$, $d = .48$). When allocating a dance motion to a speech sound (regardless of rhythm), participants were correct 67.0% of the time, which was significantly better than the chance value of one-in-two ($t(19) = 2.87$, $p ≤ 0.01$, $d = .64$). The effect sizes suggest that participants found it easier to guess the fine detail of the speech than the large scale detail of rhythm.

Stimulus Differences

When we compared the percentage of correct guesses for the different videos, some motions were clearly more guessable than others (**Figure 4**). Short 'ki' and long 'bu' were the most guessable (70%), followed by short 'bu' and staggered 'ki' (55%), with the worst performance for long 'ki' (45%) and staggered 'bu' (40%).

Figure 4. Average number of correct guesses for each stimulus. Chance values are dashed lines. Error bars +/- 1 SE.

As evident in the illustration of the dance gestures (**Figure 2**), 'ki' motions were short, jerky, upward motions, made with arms and hands, while 'bu' motions were slower, swaying, downward motions made with the whole torson and legs moving downward. This pattern fits well with the idea that within a single sensory modality, sensory alliances between different features of a signal can enhance the identity of the stimulus itself, acting as a kind of super-stimulus. For 'ki' the duration of the short rhythm was the best match for a sound typically mapped to small, jagged shapes (and here, small, jerky motions), while for 'bu' the long, smooth rhythm was the best match for a sound typically matched with large curved shapes (and here, large, smooth motions).

Individual differences across the population

Since we were interested in whether dance experience might influence how well junior dancers could guess the symbolism of these actions, we compared cumulative years of dance experience with overall task scores. **Fig-**

ure 5 shows that there was no clear relationship between years of dance experience and performance in this guessing task ($\rho(20) = .19$, non-significant). This suggests that the performance we observed for the group taps into a general system of cross-modal correspondence shared by the general community (see **Fig 1**, "Most of us"), regardless of the duration of dance experience.

However, it is also interesting to note the cluster of six participants who correctly identified all six dance motions. This group is interesting for two reasons: Firstly, participants chose not to guess the same sound twice, meaning there was a one-in-21 chance of guessing all six motions correctly. This means, from our group of 20, we should expect one participant to guess all dances correctly purely by chance. Instead, we see almost a third of our sample achieving this high score.
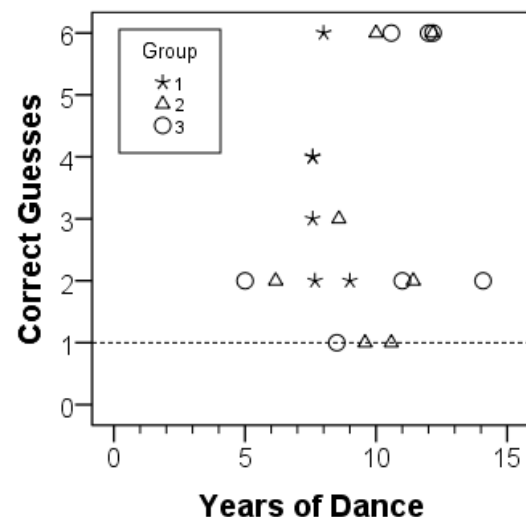
Figure 5. Years of dance experience and number of correct guesses for each participant, with different year groups shown separately. Chance value marked with dashed line.

Secondly, there is something of a performance gap between the response of these six, and the spread of responses across the rest of the group. This bimodal distribution suggests that we observing two groups – a general population with a normal spread of scores, and a separate cluster of high performers. These six could represent a cluster of high-sensory integration individuals, as proposed in the popula-

tion model of sensory congruence: these could be artists in-the-making.

4. Discussion

In this small-scale test, we observed that an experienced dancer was able to convey sensory information about meaningless speech sounds to an audience using abstract dance gestures, in such a way that performance in the guessing game was above chance. This means that something about the sound structure was encoded in the detail of our dancer's motions, despite speech being a non-standard stimulus for dance.

The fact that the dancer's audience was able to extract relevant information is evidence for shared cross-modal abstractions between the dancer and her audience, and this did not appear to be driven by experience in the medium of dance.

Furthermore, we observed that the rhythm and the identity of a sound interact in making a stimulus 'more danceable', and/or making a video more guessable: when the internal features of the sounds 'go together' the stimulus works better in tests of this kind.

Within our group of junior dancers, we observed two separate groups: the majority displayed a normal distribution scoring from one-to-four correct guesses, and a separate cluster of six perfect scores. Even in this small-scale test, this clustering provides partial support for a model in which the majority of participants share more-or-less agreement of sensory interactions, but a smaller proportion of the population may feel these sensory interactions more deeply – and these people may be the ones who actively seek out the arts (hence, being over represented in the group).

However, since all of our junior dancers were taking part in dance as a school activity, it is possible that they have chosen dance for a variety of reasons other than their own 'feeling' for sound and motion. It therefore remains to be seen whether these results are representative of a more general population (including adults and non-dancers), and whether people with higher-than-normal rates of sen-

sory integration are similarly over-represented among adult dancers.

References

Bremner, A. J., Caparos, S., Davidoff, J., Fockert, J. d., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to Westerners. *Cognition, 126*, 165-172.

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action Observation and Acquired Motor Skills: An fMRI Study with Expert Dancers. *Cerebral Cortex, 15*, 1243-1249. doi: doi:10.1093/cercor/bhi007

Cytowic, R. E., & Eagleman, D. M. (2009). *Wednesday is Indigo Blue: Discovering the brain of synesthesia*. Cambridge MA: MIT Press.

D'Onofrio, A. (2014). Phonetic Detail and Dimensionality in Sound-shape Correspondences: Refining the Bouba-Kiki Paradigm *Language and Speech, 57*(3), 367-393.

de Saussure, F. (1911:1959). *Course in general linguistics* (1959 ed.). New York: Philosophical Library.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*(1), 1-12.

Köhler, W. (1929:1947). *Gestalt psychology*. New York: Liveright Publishing.

Ludwiga, V. U., Adachid, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans *PNAS, 108*(51), 20661-20665.

Mondloch, C., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences

in young children. *Cognitive, Affective & Behavioral Neuroscience, 4*(2), 133-136.

Morton, E. W. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist, 111*, 855-869.

O'Hanlon, E., Newell, F. N., & Mitchell, K. J. (2013). Combined structural and functional imaging reveals cortical deactivations in grapheme-color synaesthesia. *Frontiers in Psychology, 4*(755), 1-13.

Oberman, L. M., & Ramachandran, V. S. (2008). Preliminary evidence for deficits in multisensory integration in autism spectrum disorders: The mirror neuron hypothesis. *Social Neuroscience, 3*(3-4), 348-355. doi: DOI: 10.1080/17470910701563681

Occelli, V., Esposito, G., Venuti, P., Arduino, G. M., & Zampini, M. (2013). The takete–maluma phenomenon in autism spectrum disorders. *Perception, 42*, 233-241. doi: doi:10.1068/p7357

Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols & J. J. Ohala (Eds.), *Sound Symbolism* (pp. 325-347): Cambridge University Press.

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*(114). doi: http://dx.doi.org/10.1016/j.jecp.2012.05.004

Pirog Revill, K., Namy, L. L., DeFife, L. C., & Nygaard, L. C. (2014). Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI. *Brain & Language, 128*, 18-24.

Ramachandran, V. S. (2003). The Emerging Mind. *Reith Lectures, BBC Radio Four*. Retrieved from http://www.bbc.co.uk/radio4/reith2003/

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia—A Window Into Perception, Thought and Language. *Journal of Consciousness Studies, 8*(123-34).

Rothen, N., & Meier, B. (2010). Higher prevalence of synaesthesia in art students. *Perception, 39*, 718-720.

Spector, F., & Maurer, D. (2013). Early sound symbolism for vowel sounds. *i-Perception, 4*, 239–241. doi: dx.doi.org/10.1068/i0535

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception and Psychophysics, 73*, 971-995.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal Infants' Sensitivity to Synaesthetic Cross-Modality Correspondences. *Psychological Science, 21*(1), 21-25.