Distribution Agreement

Signature:

Sara Marie List March 11, 2019

The sound of meaning: Physical, perceptual, and neural correlates of sound to shape mapping in natural language

By

Sara Marie List
Doctor of Philosophy
Graduate Division of Biological and Biomedical Science
Neuroscience

Krish Sathian, M.D., Ph.D.
Advisor

Lynne C. Nygaard, Ph.D.
Committee Member

Shella Keilholz, Ph.D.
Committee Member

Robert Liu, Ph.D.
Committee Member

Sam Sober, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

The sound of meaning:

Physical, perceptual, and neural correlates of sound to shape mapping in natural language

By

Sara Marie List
B.S., B.A., Emory University, 2014

Krish Sathian, M.D., Ph.D.
Advisor

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Graduate Division of Biological and Biomedical Science
Neuroscience
2019

Abstract

The sound of meaning:

Physical, perceptual, and neural correlates of sound to shape mapping in natural language

By Sara Marie List

One of the fundamental properties of language is the assumed arbitrariness of the relationship between sound and meaning. Recent behavioral, linguistic, and neuroscience studies have challenged this assumption through the demonstration of sound-symbolic crossmodal correspondences, or non-arbitrary associations between sound and meaning. Research, including psychophysical and preliminary fMRI studies, indicates that sound symbolism may result from multisensory interactions within the brain. This set of studies investigated three questions that are central to the understanding of sound-symbolic crossmodal correspondences. First, we question which physical parameters in visual and auditory stimulus space elicit sound-symbolic crossmodal correspondences. Second, we question whether the neural underpinnings of the perception of sound-symbolic crossmodal correspondences suggest that the linguistic advantages of sound symbolism are deeply ingrained in our sensory systems. Third, we investigate how these physical parameters and neural correlates are modulated by individual variation in perceptual capabilities. Using the mappings between auditory pseudowords (e.g. 'loh-moh') and visual shapes (e.g. a blob), these three ideas are investigated in typical individuals and one synesthete with psychophysical and functional neuroimaging paradigms. In our first set of studies, we compared the influence of acoustic and visual parameters on previously collected sound-shape judgments of systematically constructed auditory pseudowords and visual shapes. In our second set of studies, we conducted a neuroimaging study designed to systematically determine how auditory, visual, and multisensory neural processing contribute to the perception of sound-symbolic crossmodal correspondences using both univariate and multivariate analysis methods. In our third set of studies, we investigate the behavioral and neural evidence of variation in the perception of sound-symbolic CCs. We first explore the variation present in a group of typical individuals (n = 24) who have all completed the neural and behavioral paradigms from our second set of studies. We then assess one individual with grapheme-color synesthesia in order to explore the validity of the idea that sound-symbolic CCs arise from grounded cognition or pervasive synesthetic associations. Taken together, this research provides insights into how humans produce and interpret meaning in natural language as well as the sensory basis of associations between sound and meaning.

The sound of meaning:

Physical, perceptual, and neural correlates of sound to shape mapping in natural language

By

Sara Marie List
B.S., B.A., Emory University, 2014

Krish Sathian, M.D., Ph.D.
Advisor

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Graduate Division of Biological and Biomedical Science
Neuroscience
2019

Acknowledgements

First, I would like to express my deepest gratitude to Dr. Krish Sathian, Dr. Lynne Nygaard, and Dr. Simon Lacey, as well as Kelly McCormick, and Dr. Prasanna Karunanayaka for their help and guidance throughout this project. These individuals are responsible for my development as a scientist. I cannot thank them enough for showing me the exceptional qualities of scientists that they exemplify. For her continuous support in my scientific endeavors since I was an undergraduate and for her offering of time and space to deeply engage with the highest scholarship, I sincerely thank Dr. Lynne Nygaard. Thank you to Dr. Simon Lacey for his patience as I learned and for his encouragement to spelunk the deepest caverns of scholarly inquiry. Thanks especially to Dr. Krish Sathian, my primary mentor, who has consistently modeled a creative vision and academic excellence that has pushed me to new heights.

Thank you to Kelly McCormick for her creation of the stimuli I used and for her mentoring as I began this project. Thanks to Ji Young Kim for beginning the project with Kelly McCormick that eventually branched into my dissertation. Thanks also to all members of the Sathian lab and the Ny-Laab, especially including the undergraduates I have mentored and who have assisted with recruitment, scanning, data collection, and the frequent morale boost: Maggie Martinez, Lauren Taylor, Lokita Rajan, Tera Robison, Michelle Tippens, and Shaily Patel. Thanks to Christina Tzeng, Eva Lewandowski, Kaitlyn Matthews, and Nicole Steiner for offering encouragement and fun conversation during the last months of my graduate career.

Thank you to all of my participants, especially those of you from the Student Activity and Academic Center and those of my friends who were always available for a pilot project. I literally could not have done this project with you. Thank you for asking insightful, fun questions and reminding me why I think science is amazing.

I'd like to thank Dr. Greg Berns and Dr. Kate Pirog Revill of the Facility for Education and Research in Neuroscience (FERN) for always knowing how to massage FERN into working order. Thank you to Charlie Ferris for great conversations while scanning. Thank you especially to Randall Stilla for sharing his experiences and being an amazing instructor on neuroimaging. Thank you to Dr. Ryan Hamilton for introducing me to the world of consulting and the study of business when I wanted to see what's outside my lab. Thanks also to all of the members of Language Group for their thoughtful insights into data presentation and design. I'd like to thank all of the members of Dr. Kenneth Moberg's lab for providing a second lab home.

For their holistic guidance and advocacy throughout the process of completing a PhD, thank you to my Senior Associate Dean of Laney Graduate School, Dr. Cathy Johnson, my Director of the Graduate Division of Biological and Biomedical Sciences, Dr. Nael McCarty, my Neuroscience Program Directors past and present including Dr. Shawn Hochman, Dr. David Weinshenker, and Dr. Ellen Hess, to my Directors of Graduate Study, Dr. Victor Faundez, Dr. Sam Sober, and Dr. Malu Tansey, and the administrators Gary Longstreet and Margie Varnado. For their suggestions and feedback, thanks to my committee members, Dr. Lynne Nygaard, Dr. Shella Keilholz, Dr. Robert Liu, and Dr. Sam Sober.

Thanks to all of my past teachers and professors who were integral to fostering my interest in learning, in writing, and in neuroscience, including my first research mentor, Dr. Joe Manns, who also introduced me to my love of computer programming. Thank you to Dale Donaldson for your instruction in AP Biology and for encouraging me to follow my obsession with the mind. Thank you to Professor Christina Simons for teaching me the art of outlining, editing, and revising until the writing works. Thank you to Caneta Gentry for teaching me the importance of being assertive when it counts. Thanks to Dr. Matthew Weinschenk and Dr. Laura Otis for guiding me to reach the highest levels of problem solving and to use the incredible connections across different fields to find solutions. Thanks to all of my teachers who influenced my life and who went beyond teaching their material to help me grow into the person I am now.

For their unwavering support, thank you to my friends and family. Thanks to my friends for struggling and laughing with me through courses and qualifying exams. Thanks to my family, including my grandparents, Kathlyn and the late Fred List, Jr, Alexander and the late Marie Lynn, my parents, Marybeth and Fred List, my sisters, Abbey and Rachel List, and my partner, Edwin Corgiat. I would like to especially thank Edwin for choosing every day to be my partner all things, to believe in my strength, and to remind me that there is light in the darkness.

This work was supported by grants to Dr. Krish Sathian and Dr. Lynne Nygaard from the National Eye Institute at the NIH (R01EY025978) and the Emory University Research Council. Support to Krish Sathian from the Veterans Administration and to SML from the Laney Graduate School is also acknowledged. Funding for this project was also provided by the Laney Graduate School, the NIGMS Emory University Training Grant in Integrative Biology: Neuroscience T32, and the NEI Vision Sciences Training Program T32.

For anyone who has questions about this dissertation or my experiences as a PhD student, I encourage you to email me at smlist20@gmail.com. Finally, thank you to everyone who spent time with me and made the journey as a doctoral student worth it. All of you have encouraged me to feed my fear a suck-it sandwich. If I have not mentioned you here, it’s not for a lack of appreciation but a lack of space.

Table of Contents

Chapter 4: Synesthesia & Individual Differences in Perception of Sound-symbolic Crossmodal Correspondences

List of Figures and Tables

Chapter 2

Chapter 3

Chapter 4

Abbreviations

a	anterior
AAL	Automated Anatomical Labeling
AG	angular gyrus
AV	auditory-visual
BOLD	blood oxygenation level dependent signal
CC	crossmodal correspondence
CP	caudate putamen
d	dorsal
ERP	event-related potential
EEG	electroencephalogram
FFG	fusiform gyrus
FFT	fast Fourier transform
fMRI	functional magnetic resonance imaging
FWE	family-wise error
GLM	general linear model
HNR	harmonic to noise ratio
Ins	insula
IPS	intraparietal sulcus
IOG	inferior occipital gyrus
L	left hemisphere
LASS	language and situated simulation
med	medial
MFG	middle frontal gyrus
MNI	Montreal Neurological Institute
MOG	middle occipital gyrus
MSI	multisensory integration
MTG	middle temporal gyrus
MVPA	multivoxel pattern analysis
p	posterior
PO	parietal operculum
PPC	posterior parietal cortex
PSD	power spectral density
R	right hemisphere
RDM	representational dissimilarity matrix
RFT	random field theory
ROI	region of interest
RSA	representation similarity analysis
SMA	supplementary motor area
SMG	supramarginal gyrus
SOG	superior occipital gyrus
SPM	Statistical Parametric Mapping
STS	superior temporal sulcus

TDT	The Decoding Toolbox
TMS	transcranial magnetic stimulation
v	ventral
VWFA	visual word form area
V1	primary visual cortex

Chapter 1

Introduction

And as in uffish thought he stood,
The Jabberwock, with eyes of flame,
Came whiffling through the tulgey wood,
And burbled as it came!

One, two! One, two! And through and through
The vorpal blade went snicker-snack!
He left it dead, and with its head
He went galumphing back.

~ Lewis Carroll

1.1 Arbitrariness in Language

According to traditional linguists, a major tenet of human language is that words are arbitrarily assigned to meaning. Scholars as early as Plato (360 B.C.E./Ademollo, 2011) have discussed that nature of naming objects and assigning meaning, as Hermogenes states in *Cratylus* that he "cannot convince [him]self that there is any principle of correctness in names other than convention and agreement; any name which you give… is the right one, and if you change that and give another, the new name is as correct as the old" (Ademollo, 2011, p. 3). John Locke (1715) later insisted that the association between a word and the corresponding meaning is "perfectly arbitrary, not the consequence of a natural connexion [sic]," (Locke, 1715, p. 393) with the argument that if language were not arbitrary, humans would have developed a single common language in which the sounds of words would necessarily resemble properties of their referents. de Saussure (1959) proposed that language is the link between a conceptual understanding, which is known as the signified, and a sound-image, or the signifier. For example, language links the concept of what a tree is, the signified, with the written or spoken word “tree”, the signifier. One can evoke the image or the concept of the tree by expressing that word or sign. These linguistic signs were assumed to bear no connection or resemblance to the signified or referent. In other words, to determine the properties of the word to represent a concept, one cannot divine these properties from a necessary characteristic of the signified.

To argue against the property of arbitrariness, if words did have a non-arbitrary connection to the concepts conveyed, similar types of words and sound structures would convey equivalent meanings across different languages. However, there are copious examples of similar concepts that do not share similar sound structure or even directly translate from one language to another. The Spanish words “ser” (to be, in a permanent or fundamental sense) and “estar” (to be, in a temporary or changeable sense) have meaning encompassed by a single infinitive in English “to be”. English uses a single linguistic form that encompasses multiple meanings in another language, so there must not be a rigid connection between sound structure and meaning. Though in Spanish both words broadly translate into the English meaning “to be,” the words used to represent this concept are structurally different. This fact further indicates that there is not a one to one mapping between a word and its meaning, even within a single language. If there were a necessary connection between words and their referents, one would expect that concepts as fundamental as “to be” would be conceptually parsed in the same way across languages (e.g. all languages having only one sense of “to be” or two) and that subsequently these types of differences across languages and within a single language would not arise. One may argue that there may be a necessary and non-arbitrary connection between one meaning and word form “to be" and between the other meaning and word form "to be." Further, the extent to which different languages distinguish concepts or meanings with linguistic terms is a different question from whether or not the structure of the word is arbitrarily related to meaning. Illustrated another way, within a language there exist words known as homophones (e.g. dye versus die in English) for which the same sound structure that can be mapped onto two completely different meanings (e.g. coloring cloth versus the absence of life in the above example). This arbitrary property of language allows for a productive system of reference with an infinite number of signs to represent the link between signified and signifier (Gasser, 2004; Monaghan, Christiansen, & Fitneva, 2011; de Saussure, 1959).

Although arbitrariness is intuitive to the understanding and production of language in many contexts as noted above, there also exist many contexts in natural language in which non-arbitrariness is evident (e.g. onomatopoeia). Plato himself in his writings questions this idea of arbitrariness through the character of Socrates as he states "names ought to be given according to a natural process, and with a proper instrument, and not at our pleasure" (p. 7). This chapter began with a quote from Lewis Carroll's *Through the looking glass and what Alice found there* (2016) originally published in 1871. The question that immediately arises from this excerpt is if language is completely arbitrary, how do Carroll's nonsense words appear to convey a certain meaning? Can pseudowords be meaningful, even when they are novel and have not been paired with any meanings by conventional language or word learning? To further illustrate the meaning in this literature, consider the first two lines of the poem "Twas brillig and the slithy toves/ did gyre and gimble in the wabe". Lupyan and Casasanto (2014) provide an apt description of how both syntactic information and phonological structure in these lines allow readers to consistently attribute certain meanings to these nonsense words (Cabrera & Billman, 1996; Johnson & Goldberg, 2013). According to Lupyan and Casanto (2014), "'Slithy' is used in an adjectival frame and has phonological neighbors lithe and slimy. This word modified 'tove,' which due to its syntactic position and phonology, is likely to be a noun, and moreover, an animate being with the ability to 'gyre and gimble', which in turn connote some type of spinning and tumbling." Note that there are clear and important distinctions to be made among meanings that are conveyed by sound structures, by regularities in the sound structure, by similarities to other words in a language, and by syntactic and morphological constraints, which are beyond the scope of this dissertation. However, how meaning is conveyed in Carroll's poem serves as an illustrative piece, and through it one can conclude that pseudowords can be meaningful. Thus, there seems to be a crack in the foundational assumption of arbitrariness.

1.1.1 Sound symbolism

Beyond literary circles, modern scholars have begun to question this assumption of complete arbitrariness (Monaghan & Christiansen, 2006), though others caution that arbitrariness still describes the majority of sound-to-meaning mappings (Lupyan & Winter, 2018). Examples of potentially non-arbitrary pairings in natural languages include onomatopoeia and Japanese mimetics (e.g. "yoboyobo" meaning to become wobbly legged, usually with age), which contain characteristic sounds that suggest their meaning (Hamano, 1998).

Besides Lewis Carroll's *Jabberwocky*, there are other examples of apparently nonsensical words that are consistently interpreted as meaningful, even in the absence of syntactic cues. This phenomenon is known as sound symbolism. For example, behavioral paradigms, such as the bouba/kiki phenomenon (Ramachandran & Hubbard, 2001), a version of which was first described by Köhler (1929), illustrate that humans as young as infants at four months old consistently indicate a correspondence between the sounds in a word (e.g., bouba) and the physical shape to which that word is assigned (e.g. bouba = round, blob; Imai et al., 2015; Köhler, 1929; Maurer, Pathman, & Mondloch, 2006; Ozturk et al., 2013). These relationships extend beyond sound and shape. In the same year as Köhler, Sapir (1929) investigated the relationships between object labels and size of the object using two tables of differing sizes as stimuli. Participants were asked to choose one label, either *mal* or *mil*, for each object. Pseudowords were used as stimuli so that the participants could not use any preconceived biases from their knowledge of or practice with language to influence their decisions, and participants associated the pseudoword *mil* with the small table and *mal* with the larger table. Sapir found that judgments of the object’s size were based on which vowel was used, with /a/ from *mal* acting as a cue for a large object. These sound-symbolic associations made with pseudowords suggest that sound-symbolic properties are not constrained solely by previous language experience. Of course, even mappings made with pseudowords could be the result of language experience. Pseudowords drawn from phonesthemes illustrate that if a sound in a particular language (e.g. “gl” in the English words glitter, glisten, glimmer, etc.) is consistently associated with a certain visual feature (e.g. sparkly, shiny objects), then when listening to pseudowords, speakers of that language will use those consistencies to make judgments (e.g. to determine the meaning of a pseudoword glinka).

1.1.1.1 Cross-linguistic sound symbolism

These sound-symbolic associations translate across languages, and this phenomenon is known as cross-linguistic sound symbolism. Cross-linguistic sound symbolism refers to the ability of humans to associate certain speech sounds with certain physical phenomena. These associations can be novel, in that they originate from a language in which the person is unfamiliar, or nonsensical, in that the sounds and correlated meanings are contrived, but importantly, these associations are held across those with differing language backgrounds. For instance, native speakers of English can more quickly and accurately identify actual Japanese word/English meaning pairs than arbitrarily assigned word-meaning pairs, which suggests that there exist systematic links between semantic meaning and spoken word structure to which learners across languages are sensitive (Kantartzis, Imai, and Kita , 2011; Nygaard, Cook, & Namy, 2009). In addition, Monaghan and Christiansen (2006), using computational models have found that sound-to-meaning correspondences in language may be both systematically as well as arbitrarily organized. Further, they found using corpus analysis that non-arbitrary mappings of phonological cues to grammatical class apply across different languages, including English, Japanese, and French (Monaghan & Christiansen, 2006). Recently, a statistical examination of two-thirds of the world's languages by Blasi et al. (2016) concluded that languages often use or avoid the same sounds or sets of sounds for specific referents. All of this evidence taken together indicates that sound symbolism may not be a specific culturally-learned phenomenon, or if it is learned based on one's native language, then these associations are at the very least pervasive and consistent across a sampling of natural languages.

1.1.1.2 Sound symbolism and language learning

Sound-symbolic mappings are important for language processing and early word learning as well (Imai & Kita, 2014). For example, individuals are sensitive to sound-symbolic crossmodal associations as early as three years (Pathman & Mondloch, 2006), fourteen months (Imai et al., 2015), and even as young as four months of age (Ozturk et al., 2013). Multiple researchers have suggested that sound symbolism is important for specific word to meaning association in young children with limited vocabularies (Gasser et al., 2004; Tzeng et al., 2017).

Findings of Imai, Kita, Nagumo, & Okada (2008) support the possibility that non-arbitrary correspondences between sound and meaning may aid language acquisition. When the stimuli were sound-symbolic versus not sound-symbolic, Japanese-learning children were more accurately able to infer word meaning (Imai, Kita, Nagumo, & Okada, 2008). Then, Kantartzis, Imai, and Kita (2011) studied English-learning children listening to Japanese mimetic-like stimuli to investigate whether the child's ability to infer word meaning from the sound structure of the words was cross-linguistic in that these sound-to-meaning mappings held across languages. The findings of Imai et al. (2008) and Kantartzis et al. (2011) support the idea that listeners are cross-linguistically sensitive to and exhibit word learning benefits due to sound symbolism.

Just as sound symbolism aids young children in bootstrapping their initial word learning, sound-symbolic associations may also provide processing advantages in adults for general categorization in large vocabularies (Brand et al., 2018; Gasser et al., 2004). Findings of Nygaard, Cook, and Namy (2009) with Japanese antonym pairs provide evidence that non-arbitrary correspondences between sound and meaning persist into adulthood. Even proponents of the arbitrariness assumption, such as Gasser (2004) and Lupyan & Winter (2018), concede that iconic language that violates the principle of arbitrariness may be advantageous in situations when the vocabulary is exceptionally small, such as during acquisition of first words, or when the space of possible distinctions among signified concepts is sparse.

1.1.1.3 Why sound-symbolic properties may exist

Upon establishing the existence of sound symbolism, the second question often asked is why sound-symbolic properties may exist for which there are a wealth of explanations (Sidhu & Pexman, 2017). We will here focus on two interpretations. Both explanations, though different, are not mutually exclusive and would, like any human behavioral phenomena, be represented in the brain. However, there may be important differences in how and where these representations would occur, depending on which explanation is being considered (See Section 1.2.2).

One possible interpretation of the sound-symbolic properties of language is that of associative learning of statistical co-occurrences because human learning and the structure of human language are inextricably intertwined (Christiansen & Chater, 2008). For example, if certain sounds are culturally enforced to encourage an association with a given physical meaning, then associative learning occurs for these statistical co-occurrences, just as we learn from repeated experience that thunder usually pairs soon after lightening to mean a storm is approaching. Traditional theorists posit that language meaning is represented separately from perception, action, and emotion in the brain (Prince & Smolensky, 1997; but see Pinker, 1999). Additionally, conceptual content is perhaps encoded into abstract symbols with no relation to memories of the sensorimotor conditions that formed these concepts (Landauer & Dumais, 1997). For example, Bedny & Caramazza (2011), from their examination of the perception of action verbs in an fMRI experiment, suggest that understanding words does not require simulation of sensory-motor processes and that modality-independent neural circuits play a greater role in language comprehension, though perception may contribute at a higher-level processing stage.

Another explanation of the origins of non-arbitrary correspondences in natural language is that of embodied cognition in which neural correlates may connect the processing of sensory information from the physical world to the processing of conceptual information in language. Ramachandran and Hubbard (2001) do not dispute that sound-symbolic associations are culturally influenced but posit that there is also a neural basis to sound symbolism that is based in an extension of sensory-motor synesthesia. According to this theory, sound symbolism could arise from modality-specific representations in the brain, and Barsalou (2008) describes these modality specific representations in the linguistic domain using the framework of language and situated simulation, or LASS. These modality-specific representations could arise from associations between any number of an object's physical properties and either aspects of articulation, such as the shape of the lips upon uttering the word, or the sound of the word that represents that object. For example, because the phoneme /o/ is produced with rounded lips, an association between the production of that sound and the round shape of a balloon may be neurally represented most particularly in sensory domains (Ramachandran & Hubbard 2001; Namy & Nygaard, 2008). Of course, mappings based on statistical co-occurrence or experience with language would be neurally represented as well. If associative learning of statistical co-occurrences explains the existence of sound-symbolic properties in natural language, some cortical areas that may be important for the neural processing of sound symbolism could include left frontal areas (Bedny & Caramazza, 2011), which one may not expect in the theory of embodied cognition. Any putative mechanism for sound symbolism would be instantiated in a neural network or system.

Regarding the cortical areas that are important for a theory like embodied cognition, studies in sensory cortices have also provided evidence for a connection between meaning and sensation that is grounded in similar cortical areas (Barsalou, 2008). For example, processing textural metaphors (e.g. ‘to have a rough day’) activates somatosensory cortex in the absence of tactile stimulation (Lacey et al., 2012). Further, the left extrastriate body area, an area known for visual limb selectivity, has shown activation for body-related metaphors (e.g. ‘to give a hand’) in the absence of a visual representation of that body part (Lacey et al., 2017). The results of Lacey et al. (2012; 2017) provide additional evidence for the connections between perception and meaning and the grounding of the processing of metaphor within domain-specific sensorimotor cortical activity.

An expanding body of evidence suggests that words referring to different visual properties of objects are encoded by regions in the brain that either overlap or lie adjacent to regions that are active during the perception of those properties (Kemmerer 2009; Barsalou 2008). This theory of grounded cognition suggests that meaning in human language is grounded, or influenced by, the sensorimotor processes that allow for the perception of said meaning (Lacey et al., 2012; 2017; Pezzulo et al., 2013). If semantics is not amodal, grounded cognition may explain why systematic sound-to-meaning mappings are found. This theory allows for the possibility that the sound structure of language may have a non-arbitrary property of sound-to-meaning mapping that is cortically represented in areas associated with sensory processing, rather than areas associated with associative learning. For instance, if semantics in language is partially derived from the experience of physical sensation, perhaps certain sensory experiences, such as seeing a ball, are represented through a mental simulation of a ball in the brain that includes what the ball feels like, looks like, and sounds like when bounced. This mental simulation of the roundness of a ball could then translate into the physical production of sound with rounded lips during expression. Recent evidence in the pupillary response of individuals to words that carry a meaning of light or dark (Mathôt, Grainger, & Strijkers, 2017) found that words that are visually or auditorily presented that convey a sense of brightness decreased participants' pupil sizes and vice versa for words that convey a sense of darkness. This finding suggests that participants' semantic understanding of these words translated into an embodied image of the meaning of the word (e.g. bright or dark), which led to an autonomic pupillary expansion or contraction. Further, word meanings in event-related potential studies have found event related potentials that are feature-specific (e.g., action-related in parietal electrodes or audition-related in central electrodes) for nouns and for verbs (Popp, Trumpp, & Keifer, 2016). In this manner, embodied cognition can account for a systematic organization of sound-symbolic properties through mental simulation of sensory experience (Barsalou, 2008; although see Wilson & Golonka, 2013 for more on embodied cognition).

1.2 Auditory-Visual Multisensory Integration

Based on the theory of embodied cognition, the perception of sound symbolism may involve audition and vision or at least vision and crossmodal perceptual-motor mappings that transcend audition (Namy & Nygaard, 2008). Also, the pervasiveness of cross-linguistic sound symbolism leads to the question of whether sound symbolism may be a natural by-product of the connections between perception and language in the brain (Sidhu & Pexman, 2017). Therefore, a general overview of auditory-visual multisensory integration is important to the understanding of this work.

Each of the sensory modalities (e.g. auditory, visual, or somatosensory) is elicited in the brain via different sensory inputs (e.g. a sound wave from a strumming guitar, an image of a Mediterranean beach, a patch of sand rubbing against the skin). Because these inputs and our perception of them seem so different, it made sense historically to study each modality separately. Increasingly in the last few decades, multisensory research has burgeoned as a field, and this research has even led to the suggestion that all of the neocortex, including areas traditionally defined as primary sensory cortices, is multisensory (Ghazanfar & Schroeder, 2006).

In a strict sense, multisensory integration (MSI) is defined as the neural processes by which unisensory signals are combined to create a multisensory response, either neural or behavioral, that differs significantly from the responses evoked from the stimuli in the unisensory components (Colonius & Diederich, 2017). The traditional definition of MSI arose from the seminal work of Meredith & Stein (1983; Stein et al., 1989) from behavioral measures and neurophysiological recordings in the superior colliculi of cats, in which they demonstrated that adding an auditory or somatosensory cue that is proximal in time or in space to a visual stimulus input leads to a superadditive neural response. Moreover, that superadditive response is behaviorally relevant in that it coincides with an increased performance on behavioral measures of orienting and moving toward the stimulus (Stein et al., 1989).

Since these initial groundbreaking studies, the field of multisensory integration has expanded significantly, so much that Colonius & Diederich (2017) caution that MSI is very broadly defined as “a generic overarching term describing processing involving more than one sensory modality but not necessarily specifying the exact nature of the interaction between them.” Perhaps one reason that this phrase is loosely defined is that neuroscience has only begun to adequately parse the functional and mechanistic roles of primary sensory areas in somatosensation, vision, and audition, let alone to describe how any of these senses meaningfully interact to create a continuous consciousness or coherent percept. Because audition and vision are the two senses that are most pertinent to sound symbolism, we will focus on auditory-visual MSI, specifically. A more specific definition of MSI may only include the particular interactions between different sensory systems as opposed the mappings or associations that describe sound-symbolic crossmodal correspondences, and although the broad definition of MSI includes sound-symbolic crossmodal correspondences and phenomena like synesthesia, it also provides unique challenges to the study of MSI as a whole by being less precise and making the convergence of findings across studies more difficult.

1.2.1 Guiding principles of multisensory integration

Despite the broad nature of the definition of MSI, research has yielded a common framework of principles to describe it. Based on Stein et al. (1989) the neural responses to MSI seem to break into three guiding principles including that the multisensory neural response tends to be enhanced when stimuli in different modalities come from the same location, when they occur at the same time, and when at least one of the two stimuli is only weakly effective on its own in eliciting an excitatory neural response (Holmes & Spence, 2005). These three principles, known as the temporal rule, the spatial rule, and the inverse effectiveness rule, have provided an excellent framework with which to study MSI.

Multiple computational models have been developed in order to understand auditory-visual MSI (Colonius & Diederich, 2017; Ernst & Banks, 2002; Billock & Havig et al., 2018). One such model known as the Bayesian estimator model, considers the brain as an estimator of uncertainty in the environment, wherein the uncertainty of sensory inputs is minimized by combining multiple, independent estimates of the environment (Ernst & Banks, 2002; Fetsch et al., 2013; Knill and Pouget, 2004). The principles by which the Bayesian estimator model was developed clearly illustrate some of the most important principles of MSI. For example, multisensory inputs can provide complementary or competing measurements of the same event, thereby decreasing or increasing the perception of uncertainty in the environment, while within-modality cues may be equally reliable or unreliable because it is more difficult for two cues of the same modality to have differing levels of accuracy in measuring the external world. For example, in a situation in which visual information is degraded, such as during twilight, within-modality parsing of a visual input stream includes sets of information such as shape and texture that are equally degraded, but adding an auditory cue to a visual cue, such as a bird chirping as it takes flight, increases the strength of the resulting multisensory response much more than multiple small or degraded visual inputs. In addition, inputs from different modalities about certain events are often predictive of one another in the natural environment and are correlated crossmodally such that one can predict the event even if one sensory cue is received before the other (Parise et al., 2012; See next section 1.2.1 Crossmodal correspondences). In contrast, within-modality sensory cues often have a similar timing. In the Bayesian model of MSI, crossmodal and within modal pairs of sensory cues can be considered different mechanisms for reducing uncertainty about an event.

The simplicity of the three guiding principles of MSI is not meant to suggest that competing evidence has not been found regarding the nature of these principles and where they arise. In fact, the spatial rule has been shown to be dependent on task requirements and has been modified to the principle of spatial correspondence, which states that inputs are more likely to be integrated when they overlap in spatial location (Cappe et al., 2012; Girard et al., 2010; Sperdin et al., 2010). In addition, the temporal rule is dependent not only on the task but also on the stimulus type (Stevenson & Wallace, 2013; van Atteveldt et al., 2007). The inverse effectiveness rule is context dependent by definition in that it predicts that a primary factor in the strength of integration effects is the intensity of the unisensory stimuli (Stein & Meredith, 1993).

MSI is clearly not a uniform and rigid function but one that is strongly adaptive to the environment. Thanks to the diverse evidence describing how and where MSI functions in the brain (See Section 1.2.2), research points to an implicit fourth guiding principle of MSI: that it is flexible and context dependent. In the Bayesian model of MSI, the idea that context is important is not new, especially in the determination of which sensory modality is most appropriate to derive information (Welch & Warren, 1980). For example, when available, visual stimuli would generally be more informative for judgments of spatial location, and auditory stimuli would capture temporal information well, so either modality could dominate the perceptual framework depending on the context. There are many contexts that are important for MSI, including the properties and modalities of the stimuli and the goal of the task, and from one moment to the next, each of these contexts interact in order to prompt the neural framework to adapt to perform MSI (van Atteveldt et al., 2014).

If we assume that MSI is necessary to create our conscious experience, then perhaps the question of whether a neuron in primary auditory cortex also receives inputs from other modalities as opposed to being specialized for one modality is not a question of the existence of the inputs but of their degree or strength. Perhaps we should not consider whether or not sensory cortices are specialized but to what degree they are tuned to one modality over others, which is proposed in the metamodal brain hypothesis (Pascual-Leone & Hamilton, 2001). We can then use multisensory paradigms to explore how the brain adapts to take input from multiple modalities in order to construct our experience of the world (Lacey et al., 2009). Multisensory paradigms, as opposed to unisensory paradigms, provide information on how and where the segregated receptor surfaces and initial input pathways converge. We here discuss MSI as a potential mechanism by which sound-symbolic associations are interpreted and processed, and by studying MSI in the context of sound-symbolic associations, we hope to uncover new insights into the interactions between auditory-visual MSI and language.

1.2.2 The neural architecture of multisensory integration

While the operation of the guiding principles of MSI was investigated initially in the superior colliculi, context-dependent results have been found across multiple neuroanatomical substrates (van Atteveldt et al., 2014). Different sources of multisensory influence on primary sensory cortices have been proposed, such as direct ascending inputs from nonspecific thalamic regions (Hackett et al., 2007; Schroeder et al., 2003), the crossing of neuronal projections from sensory cortices to different modalities (Falchier et al., 2002; Rockland & Ojima, 2003), and even indirect feedback from specific multisensory regions in the cortex (Smiley et al., 2007). Many researchers have described how multiple neural architectures for MSI may functionally complement one another (Driver & Noesselt, 2008; Ghazanfar & Schroeder, 2006; van Atteveldt et al., 2013; 2014; Werner & Noppeney, 2010a).

Because we are studying sound-symbolic CCs, a process that may involve MSI, with functional magnetic resonance imaging (fMRI) in humans, we will primarily be able to focus on cerebral cortical areas of the brain with our methods. fMRI is essentially a correlational method in which differences in blood-oxygenation level dependent (BOLD) under different experimental conditions allows researchers to spatially localize areas in the brain that may be important for processing particular experimental conditions or stimuli. In contrast to the correlational inferences of fMRI, transcranial magnetic stimulation (TMS) studies have been particularly informative in delineating the cortical areas involved in auditory-visual MSI because, unlike fMRI, TMS allows for the possibility of making causal inferences through the application of strong, localized magnetic fields across the skull that can excite or inhibit those underlying regions of cortex during the experimental conditions (Yau et al., 2015). What is known of the cortical neural underpinnings of auditory-visual MSI and crossmodal processing can be broadly classified into three categories: evidence for the narrow definition of MSI in primary sensory cortices, evidence for the broader definition of MSI in posterior parietal areas, and evidence for MSI as it relates to speech processing in superior temporal sulcus.

1.2.2.1 Multisensory integration as narrowly defined by the primary sensory cortices

Recall that multisensory integration (MSI) is defined as the neural processes by which unisensory signals are combined to create a multisensory response, either neural or behavioral, that differs significantly from the responses evoked from the stimuli in the unisensory components (Colonius & Diederich, 2017). Following this narrow definition of MSI, the first areas to investigate auditory-visual MSI should be the primary sensory cortices. Traditional models of hierarchical organization of the sensory cortices state that sensory information processing for each modality is strictly segregated into modality-specific areas, known as primary sensory cortical areas, before any multisensory integration occurs, but this organization has been questioned over the last decade by researchers who have demonstrated multisensory responses in these so-called primary sensory areas (Ghazanfar & Schroeder, 2006; Driver & Noesselt, 2008). The two broad types of multisensory responses that have been investigated in primary cortices include first, responses in which the primary sensory areas modulate the response of an off-target modality (e.g. auditory response modulation in visual cortex during the presence of an auditory stimulus) and second, the modulation of primary sensory cortex responses by off-target sensory inputs (e.g. visual response modulation from auditory cortex during presence of auditory stimulus).

Early processing stages of vision include the early visual pathway, which is known to represent simple visual shape information like orientation and curvature (Connor, Brincat, & Pasupathy, 2007). However, these regions have also been known to respond to tactile stimuli that convey orientation in the absence of visual stimulation (Merabet et al., 2008; Sathian et al., 1997; Zangaladze et al., 1999), which has led to the theory that visual cortex is not limited to vision but is a general spatial and shape information processing center (Pascual-Leone & Hamilton, 2001). For example, TMS in visual cortex impairs the tactile discrimination of grating patterns that vary in orientation (Zangaladze et al., 1999). Further, the involvement of visual areas in tactile processing goes beyond simple primary visual cortex and shape features and extends into the recognition of objects in the lateral occipital cortex (LOC) and the IPS (Lacey et al., 2009) and the processing of visual motion processing in the middle temporal area (MT; Basso et al., 2012).

Similar to visual cortex, auditory cortex also contributes to the processing of off-target modalities. Off-target modalities include the sensory modalities that do not correspond to the modality of the sensory stimulus (e.g. audition and vision are off-target modalities for a tactile stimulus). Tactile stimulation can evoke responses in auditory cortex (Nordmark, Pruszynski, Johansson & 2012). Because auditory cortex is specialized for temporal information such as frequency and timing, information that encodes frequency and timing in the tactile domain may be modulated by auditory cortex, and indeed excitatory TMS over auditory cortex improves performance on a task involving the discrimination of tactile frequency (Ladeira, 2011). The evidence for crossmodal recruitment of sensory cortices includes visual cortex for shape, orientation, or object recognition, auditory cortex for frequency and timing information, and even somatosensory cortex for body-related information (Pitcher et al., 2008; Lacey, Stilla, & Sathian, 2012).

In addition to crossmodal recruitment, off-target responses in other modalities have been shown to modulate primary sensory cortices. A TMS pulse delivered to the primary visual cortex can evoke a visual phenomenon known as phosphenes in certain parts of the visual field, and the probability of producing phosphenes in a participant's visual field increases with the excitability state of the visual cortex (Merabet, Theoret, & Pascual-Leone, 2003; Ray et al., 1998). Subthreshold (i.e. imperceptible) tactile stimulation of the right hand increases phosphene occurrence in the right visual field when visual cortex is exposed to excitatory TMS within the temporal window of 60 ms post-tactile stimulation (Ramos-Estebanez et al., 2007). Induction of phosphenes in the presence of auditory stimuli has also been shown to modulate visual cortex excitability and can depend on the characteristics of the sounds, such that looming sounds influence the visual cortex excitability more than receding or stationary sounds (Romei et al., 2009). Romei et al.'s (2009) results suggest that the modulation of visual cortex by auditory cortex is pre-perceptual, potentially automatic, and specific to certain auditory stimulus features.

Numerous fMRI and EEG studies have illustrated auditory-visual MSI within primary visual cortex. Laurienti et al. (2002) found that there was deactivation in visual cortices in response to sound but not in the presence of visual stimuli, although because the task required attention only on the visual stimuli, others have argued that the effect may have been based on goal-directed attention to vision (Mozolic et al., 2008). Further research indicates increased functional coupling between primary visual and auditory cortices under synchronous auditory-visual stimulation (Lewis & Noppeney, 2010; Tyll et al. 2013), which may be further mediated by thalamic activity (Bonath et al., 2013). Auditory stimulation has been shown to activate visual cortices in contexts with prior multisensory experiences (e.g. the sounds of a party) or in the presence of concurrent visual stimuli (Matusz, Wallace & Murray, 2017; Zagenehpour & Zatorre, 2010). Watkins et al. (2007) illustrated that perception of the flash-beep illusion (Shams et al., 2000) is predicted by activity in primary visual cortex. The category of natural sounds that was heard by participants can be decoded from data extracted from primary visual cortex (Vetter et al., 2014), and this collection of findings provides strong evidence for a link between crossmodal responses in visual cortex and perception. One caveat of many fMRI studies illustrating the multisensory response of visual cortex in the presence of auditory stimulation or vice versa is that the response is often not supra-additive or may in fact be smaller than the original visual response (Martuzzi et al., 2007). Further, the unisensory response can follow a different temporal pattern than that of the multisensory response (Martuzzi et al., 2007; Narsude et al., 2015), and both of these findings call into question two of the basic principles of MSI (Noppeney, 2012). fMRI and EEG may expand our knowledge of auditory-visual MSI in humans beyond the basic principles outlined in studies of cats, and thus fMRI provides an essential extension of the approaches available to identify brain regions and neural correlates important for MSI.

1.2.2.2 Multisensory integration as broadly defined in the posterior parietal cortex

While the narrow definition of MSI allows for the investigation of the fundamental ways that sensory inputs combine in the brain, there are other complex multisensory processes that do not fit easily into such strict terms. Considering MSI from a broader perspective as a process involving more than one sensory modality, studies indicate that the posterior parietal cortex (PPC) is an important neuroanatomical structure for many secondary functions. Multisensory functions of the PPC include attentional and stimulus binding effects and effects that may include the regulation of crossmodal interactions from sensory cortices. Excitatory TMS applied over the right PPC has been shown to speed reaction times for detecting auditory, visual, and auditory-visual multisensory stimuli (Bolognini et al., 2010). Inhibitory TMS applied to the angular gyrus (AG) and the supramarginal gyrus (SMG) impairs individuals' abilities to locate a tactile target in the presence of non-informative visual and tactile distractors (Chambers et al., 2007), indicating that these areas are needed to impede reflexive but non-informative or distracting attentional shifts presented across modalities. Further, inhibitory TMS applied to right AG impairs the ability to bind auditory and visual cues delivered within short windows of time (Kamke et al., 2012).

In addition to regulating attentional shifts, PPC has also been implicated as a region that receives unisensory inputs from different sensory cortices and is thus important for higher-order sensory association in traditional models of MSI that include hierarchical sensory processing (Ghanzafar & Schroeder, 2006; Macaluso & Driver, 2005). PPC may further modulate primary sensory cortices through feedback projections that influence processing in primary sensory areas (Macaluso & Driver, 2005). Much of the work that supports PPC as a modulator of primary sensory cortices has been done on vision and touch (Pasalar, Ro, & Beauchamp, 2010; Harrar et al., 2018 for review in blind individuals). Visual stimulation can serve to recalibrate tactile sensitivity (Ro et al., 2004). Auditory stimulation located near the hand can enhance touch, but these effects disappear in the presence of inhibitory TMS of the PPC (Serino, Canzoneri, & Avenanti, 2011).

PPC includes a large area of the brain that may function to generally mediate the binding of sensory information across modalities, but smaller subregions of PPC, such as the intraparietal sulcus (IPS) likely serve to focus attention on particular modalities over others (Chambers et al., 2004). Additionally, our lab has previously shown that PPC, particularly intraparietal sulcus

(IPS) and SMG are differentially activated during the processing of incongruent sound-symbolic crossmodal correspondences (McCormick et al., 2018). These findings suggest that PPC is important for attentional shifts and for modulation of sensory cortices that allow for multisensory binding across space and time.

1.2.2.3 Multisensory integration as related to speech in the superior temporal sulcus

If we broaden the definition of MSI in the previous section, we can use that definition to investigate the role of MSI in processes that involve more than one sense. For example, because we use both auditory and visual sensory information for communication, these senses necessarily interact during speech perception. Another cortical area that has appeared consistently in the literature and may play a significant role in multisensory integration, during speech perception especially, is the superior temporal sulcus (STS). Studies implicating the STS have often used the McGurk effect as a useful experimental paradigm. The McGurk effect is a phenomenon that occurs due to this interaction between audition and vision during speech perception in which the auditory syllable that a participant perceives depends on whether a visual video of a speaker producing a sound is the same as the auditory syllable (congruent; lips showing movement for “ba” and sound “ba”) or is different from the auditory syllable (incongruent; lips showing movement for “ga” and sound “ba”; McGurk & MacDonald, 1976). Incongruent auditory and visual stimuli often result in the altered perception of syllables that are not presented in either modality. For example, a participant may perceive the sound “da” for the above incongruency wherein the lips showed movement for “ga” with the presented sound “ba”. The STS has been identified in neuroimaging studies as an important region for the processing of auditory-visual speech (Beauchamp, 2005a), and a homologous area in macaque monkeys has been shown to have neurons that are tuned for auditory and visual input that may be related to communication (Perrodin et al., 2014). Using the McGurk effect, TMS studies have established the contribution of the STS to auditory-visual MSI in speech perception. Beauchamp, Nath, and Pasalar (2010) found that applying inhibitory TMS to STS elicited fewer occurrences of the McGurk effect compared to trials with TMS applied to another control area or trials with sham TMS applied to STS. Interestingly, this procedure only affected instances of the McGurk illusion during incongruent trials and did not affect perception of syllables when auditory and visual information was consistent, and the study further established that a specific time range of 100 ms post-stimulus onset is required for the delivery of TMS on the STS to be effective (Beauchamp, Nath, & Pasalar, 2010). Multiple regions may be important for auditory-visual MSI in speech processing, and the results of Beauchamp, Nath, and Pasalar were replicated and extended to apply to TMS on PPC (Marques et al., 2014). Neuroimaging studies have also indicated a potential interaction between STS and PPC in that activation in both areas was correlated with participants' perception of the McGurk effect (Kilian-Hütten et al., 2011).

The STS has also been implicated in auditory-visual MSI outside of speech perception. STS may support auditory-visual MSI without the presence of biological movement, as was shown by Bolognini et al. (2011). Inhibitory TMS on STS decreased the sound-induced fission or flash-beep illusion (Shams et al., 2000), in which a single flash appears to flicker with the presence of multiple auditory tones, and conversely, excitatory TMS on STS increased susceptibility to the sound-induced fission illusion (Bolognini et al., 2011). For the sound-induced fission illusion, STS may also interact with visual cortex, which has been shown to modulate the illusion via TMS in a polarity-dependent manner (Watkins et al., 2007). Although previous neuroimaging and TMS studies have implicated left posterior STS in perceiving speech sounds (Möttönen et al. 2006; Venezia et al., 2017), in auditory-visual speech perception (e.g. lip reading; Calvert et al., 2003), and in binding auditory and visual inputs (Bolognini et al., 2011), studies specific to sound-symbolic crossmodal correspondences have not all reported this area in multimodal processing in this context. Revill et al. (2014) and Kanero et al. (2014) have found evidence for the involvement of the STS during the processing for sound symbolism using synonym-antonym pairs (Revill et al., 2014) and Japanese mimetics (Kanero et al., 2014), but more recently, McCormick et al. (2018) did not find evidence for the involvement of the STS while processing sound-symbolic crossmodal correspondences (See Section 1.2.3). Nevertheless, STS is an important cortical area to consider in the investigation of a phenomenon that is based in language like sound-symbolic crossmodal correspondences.

1.2.3 Crossmodal correspondences

Crossmodal correspondences (CCs) are near-universally experienced associations between apparently arbitrary stimulus features in different senses, the behavioral testing of which have been used to demonstrate MSI in its broader definition (Spence, 2011). For example, large and small size in the visual domain are consistently associated with low- and high- pitched sounds, respectively, (Gallace & Spence, 2006; Evans & Treisman, 2010), and high and low elevation are associated with high and low pitch in the auditory domain, respectively (Ben-Artzi & Marks, 1995; Lacey et al., 2016; Jamal et al. 2017). CCs often occur between stimulus properties that are correlated (e.g. high-pitched sounds and small-sized animals), which can lead to more efficient processing of sensory information. In fact, one of the mechanistic explanations of CCs is that of statistical co-occurrence, in which this environmental correlation between two stimulus properties (e.g. small mice make high-pitched squeaks and large elephants make low-pitched rumbles) encourages associations in perception (Spence, 2011). There is evidence that even non-human animals can perceive crossmodal correspondences, as is the case for chimpanzees and the CC for pitch and brightness (Ludwig et al., 2011; See Ratcliffe, Taylor, & Reby, 2016 for review), suggesting that the mechanisms behind CCs are evolutionarily conserved or converged. CCs may underlie certain aspects of natural non-human animal interactions in the wild. For example, most fundamental frequencies of animal vocalizations are inversely related to body size, and some species make unusually low sounds to exaggerate the perception of their size for territorial disputes and mate selection (Fitch, 2000). Crossmodal interactions in pitch and size or pitch and brightness can also affect response times in speeded classification tasks (Melara & Marks, 1990a; Melara & O'Brian, 1987). Because the addition of congruent sensory experiences (e.g. low pitch and low brightness) can increase reaction times compared to incongruent (e.g. high pitch and low brightness) or unimodal (e.g. low pitch alone) conditions, these associations may have a neural basis through cross-sensory mappings and have been used in the study of MSI (Melara & Marks, 1990a; Melara & O'Brian, 1987; See Chapter 3).

1.2.3.1 Sound-symbolic crossmodal correspondences and language processing

Sound-symbolic associations like the kiki/bouba effect are a specific type of crossmodal correspondence. Sound-symbolic CCs are unique to other types of CCs in that they involve speech sounds and may therefore be used to gain insight both in MSI and in the processing of language. Multiple studies have attempted to determine whether there may be a processing speed advantage to sound-symbolic associations as well as the acquisition and accuracy advantages described earlier (Chen et al., 2016b; Parise & Spence, 2009; Parise & Spence, 2012; See Section 1.1.1). Parise and Spence (2012) found a stronger coupling of matched vs.

mismatched stimuli using an implicit association task and provided evidence that sound-symbolic congruency can facilitate sensory-perceptual processing. Crossmodal correspondences therefore appear to support crosstalk, at least between visual and auditory domains.

1.3 Physical Parameters of Sound-symbolic Crossmodal Correspondences

CCs may be reflected within the structure of language in the form of cross-linguistic sound symbolism. The existence of cross-linguistic sound symbolism has been documented extensively in behavioral paradigms (Kantartzis, Imai, and Kita , 2011; Monaghan & Christiansen, 2006; Nygaard, Cook, & Namy, 2009), and neural correlates of these findings have been suggested in neuroimaging studies of sound symbolism (Kanero et al. 2014; Revill et al., 2014; McCormick et al., 2018). However, few studies have examined what physical aspects of the sound structure of language evoke associations with particular visual attributes (Nuckolls, 1999). For example, it is unclear, whether it is sensitivity to the overall acoustic form of particular linguistic segments, the manner in which speech sounds are produced (e.g., lip rounding), or a combination of these factors that underlie associations with meaning (Ramachandran & Hubbard, 2001; Spence 2011). Nielsen and Rendall (2011) attempted to describe word attributes that contribute to sound-symbolic properties using nonsense words and found that both vowels and consonants have an effect on whether or not a nonsense word is non-arbitrarily paired with one visual stimulus or another. Our lab and others have extended the findings of Nielsen and Rendall (2011) into speakers of French to establish the cross-linguistic nature of these sound-to-meaning associations (Fort & Peperkamp, 2014) and in a systematically developed set of pseudowords to establish the phonetic categories that influence these associations (McCormick et al., 2015). In his work examining the sound-symbolic associations of pseudowords and size, Sapir (1929) proposed that either certain vowels have greater acoustic volume than others in the auditory domain or the large spatial relationships between part of the articulatory apparatus during sound production (i.e., tongue position and resonance cavity) is symbolic of a larger reference. For example, the pseudoword “mal” may be perceived as larger than the pseudoword “mil” because the configuration of the mouth and throat is larger when producing the “a” versus the “i” sounds.

As evidenced by previous sections (See Section 1.1), sound symbolism and its relation to meaning can quickly become complex by asking what does “meaning” refer to. We simplify this question by examining the specific relationship between pseudowords and shapes on the scale of roundedness to pointedness using the sound-symbolic CC between sound and shape. By studying this particular sound-symbolic CC, we gain clarity over the exact relationship between particular sounds and particular meanings, at least for this context, which through inference may shed light on the nature of the perceptual processes behind sound to word associations. Although others have examined visual (Chen et al., 2016a) or acoustic properties (Knoeferle et al., 2017) that are important for sound-symbolic CCs, we used representational similarity analysis (See Chapter 1, Section 1.6) to examine the relationships between physical stimulus parameters and perceptual ratings that varied on dimensions of roundedness and pointedness, for a range of auditory pseudowords and visual shapes (See Chapter 2; List et al., BioRxiv). The systematically developed and large stimulus set (90 shapes and 537 pseudowords) of McCormick and colleagues is more extensive than those of others who have explored similar questions (See Appendix 1 & 2; McCormick et al., 2015; McCormick et al., BioRxiv; McCormick et al., unpublished data). This advantage allowed us to explore a spectrum of rounded to pointed shapes and pseudowords. In our study, we compared the influence of acoustic and visual parameters on sound-shape judgments. First, McCormick and colleagues systematically constructed visual shapes to assess the role of visual features in shape judgments (McCormick et al., unpublished data), and second, McCormick et al., (2015) constructed auditory pseudowords to assess the role of acoustic features in sound-shape symbolic judgments (McCormick et al., BioRxiv). Our study detailed in Chapter 2 (List et al., BioRxiv) is also the first to investigate the similarities in perceptual judgments of rounded/pointedness for shapes compared to pseudowords and to further use these similarities to determine which visual and acoustic properties are important for sound-symbolic CCs both within and across modalities.

1.4 Neural Underpinnings of Sound-symbolic Crossmodal Correspondences

In recent years, a few neuroimaging studies have examined the neural correlates of sound symbolism, and these studies both challenge the assumption of arbitrariness and provide evidence for how and where sound-symbolic CCs may be crossmodally mapped in the brain (Kovic et al., 2010; Revill, Namy, Defife, & Nygaard, 2014; McCormick et al., 2018). For sound-symbolic CCs in particular, fMRI studies by Peiffer-Smadja (2010; 2019) have indicated a difference in processing of matching and mismatching shape-pseudoword pairs (bouba/kiki effect) within the lateral occipital cortex. Kovic, Plunkett, and Westermann (2010) found that participants were faster to identify novel objects with sound-symbolic rather than arbitrary labels and that this behavioral finding was correlated with early negative ERP (event-related potential) components within 200 ms of object presentation. These findings indicated a potential neural mechanism linking auditory and visual features that supports sound-symbolic associations (Kovic et al., 2010). Others have implicated MSI areas in the temporal and parietal lobes, such as the left superior parietal cortex (SPC), the superior temporal cortex (STC), or the superior temporal gyrus (STG), in crossmodal correspondences related to sound symbolism (Arata et al., 2010; Kanero et al., 2014; Revill et al., 2014). These regions may be implicated in functions that allow perceptual attributes of the referent to be correlated with word structure. Our lab has previously found that attentional mechanisms may also play an important role in the processing of sound-symbolic CCs (McCormick et al., 2018).

In order to investigate the neural underpinnings of sound-symbolic CCs, we conducted an fMRI study designed to systematically determine how auditory, visual, and multisensory neural processing contribute to the perception of sound-symbolic crossmodal correspondences using both univariate and multivariate analysis methods (See Chapter 1, Section 1.6; See Chapter 3;). Specifically, we examined whether differential activation in a key set of cerebral cortical areas reliably predicted sensitivity to sound-symbolic crossmodal correspondences. In our study (See Chapter 3), we compare the relative influence of multiple cerebral cortical areas during auditory and visual presentation of the component stimuli that elicit sound-symbolic crossmodal correspondences. First, we confirmed participants' sensitivity to sound-symbolic crossmodal correspondences with two independent behavioral tasks, one inside and one outside the scanner. Then, we identified cerebral cortical regions whose preferential activation is correlated with the multisensory presentation of sound-symbolic crossmodal correspondences. Then, we investigated how the unisensory presentation of sound symbolic stimuli, shapes or pseudowords, respectively, was correlated with perceptual judgments of the roundedness/pointedness of these stimuli. The perception of sound symbolic crossmodal correspondences may involve the distinction of roundedness/pointedness in both visual and auditory modalities. In order to investigate how the distinction of rounded or pointed in a single modality may contribute to the multisensory percept of sound symbolic crossmodal correspondences, we compared cerebral cortical areas that showed greater activation for pointed over rounded (and rounded over pointed) shapes with those areas that showed greater activations for pointed over rounded pseudowords. Finally, we performed multivariate analyses methods (See Chapter 1, Section 1.6) in these regions on BOLD (blood-oxygenation level dependent) signal data collected during the perception of visual shapes and separately during the perception of auditory pseudowords to assess the roles of visual perception and auditory perception, respectively, in sound-symbolic crossmodal correspondences.

1.5 Synesthesia and Individual Perceptual Differences

One idea that speaks to the mechanistic origins of sound-symbolic CCs suggests that listeners do not simply mentally simulate to make sound-symbolic associations under the framework of grounded cognition (i.e. using auditory and visual imagery), but may go a step further and actually experience the symbolic cues in multiple modalities, such as vision and hearing (Barsalou et al., 2008; Ramachandran & Hubbard, 2001). Synesthesia is a phenomenon in which a person responds to a primary stimulus in one modality, or the “inducer”, with a secondary, synesthetic experience in a separate modality, called the “concurrent” (Eagleman et al., 2007). Crossmodal correspondences were once thought to be a form of synesthesia and were referred to as ‘synesthetic’ under the assumption that all individuals were on a spectrum of synesthesia (Ramachandran & Hubbard, 2001; Chen et al., 2016b), an assumption which has since been called into question (Deroy & Spence, 2013). Cross-sensory mappings, including sound-symbolic CCs, have been studied extensively in individuals with synesthesia (Lacey et al., 2016), who report experiences such as “seeing” sounds specific colors or experiencing numbers with specific personalities (See Chapter 4). As an alternative to grounded cognition and the spectrum of synesthesia, individuals may perform associative learning of statistical co-occurrences (Nygaard & Namy, 2008), separate from the phenomenon of synesthesia.

Whether the mechanisms behind synesthesia are separate from the mechanisms of sound-symbolic CCs or not, synesthesia is one prominent example of variation in perception. If a fundamental property of the neuronal mechanisms of MSI is that they are context dependent (See Chapter 1, Section 1.2), then alternatively we can envision synesthesia as a different internal context under which MSI can occur. Once we determine how MSI may be important for the perception of sound-symbolic CCs at the group level, we may ask how individual differences in perception of sound-symbolic CCs can be explained by variation in these neuronal mechanisms of MSI and vice versa. We are interested in how an individual’s perceptual experiences vary and how those variations may shape the characteristics of their neural response to multisensory stimuli. MSI can easily be influenced by individual variation. If we recall the computational model of MSI, the Bayesian estimator, individual variation in MSI arises from different individuals take different measurements of the same stimuli, based on their sensory receptors, their attentive state, or their past experiences. For example, increased sensitivity to auditory-visual synchrony has been found from perceptual or musical training (Lee & Noppeney, 2011a; Powers et al., 2009). Powers et al., 2009 used a task in which participants judged auditory-visual simultaneity with feedback and posit that this feedback fine-tuned participants’ temporal sensitivity through multisensory learning. In addition, the default modality for sensory dominance or bias has been shown to vary across individuals (Romei et al., 2013), and may also be sensitive to training (Sandhu & Dyson, 2012). The effects of individual experience in MSI can happen as a result of development or rapidly, as have been shown in recalibration experiments in the domains of time, space, and speech content (Hillock-Dunn & Wallace, 2012; Stevenson et al., 2018; Van der Burg et al., 2013; Vroomen & Baart, 2012).

Building upon experiments from Chapter 3, we here investigate the behavioral and neural evidence of variation in the perception of sound-symbolic CCs (See Chapter 4). We first explore the variation present in a group of typical individuals (n = 24) who have all completed the neural and behavioral paradigms detailed in Chapter 3, Section 3.3. We then assess one individual with grapheme-color synesthesia in order to explore the validity of the idea that sound-symbolic CCs arise from grounded cognition or pervasive synesthetic associations (Barsalou et al., 2008; Ramachandran & Hubbard, 2001).

1.6 Univariate and Multivariate Functional Magnetic Resonance Imaging (fMRI)

The proceeding chapters of this dissertation rely on a basic understanding of our methods, which include the standard univariate analytical methods for functional magnetic resonance imaging (fMRI) and the newer multivariate analytical approaches that have enabled us to make the scientific insights we have gained in these investigations.

The challenge of quantifying multisensory and crossmodal responses in human fMRI is that these brain imaging methods involve determining the appropriate statistical criterion for identifying multisensory interactions throughout the brain (Laurienti et al., 2005; Noppeney, 2012). Translating the established principles of MSI, which were defined on the basis of single-unit recordings in cats (Stein & Meredith, 1993), to behavioral and population-level neural responses in humans is no small task. Researchers particularly debate over how to differentiate among and interpret supra-additive and sub-additive nonlinear responses, which are multisensory responses that are greater or less than the sum of the unisensory responses, respectively (Murray et al., 2018; Noppeney, 2012; Stevenson et al., 2014). For fMRI, there is still no consensus on the best statistical criterion to identify multisensory brain activity, but certain individuals recommend different methods from using a non-linearity criterion (Calvert, 2001), to comparing multisensory responses to the mean or maxima of unisensory responses (Beauchamp, 2005), to using the criterion of inverse effectiveness (James & Stevenson, 2011). To circumvent and mitigate this difficulty in determining a truly multisensory response with fMRI, we have used a multisensory fMRI session with behavioral proxy to indicate integration of the auditory-visual stimuli. We have also chosen to apply a combination of univariate and multivariate analytical approaches to our work.

1.6.1 Standard univariate approaches

Standard univariate methods of fMRI data analysis seek to identify specific brain regions that are responsible for, or at least associated with, specific mental and behavioral activities (Mur, Bandettini, & Kriegeskorte, 2009). While this approach increases statistical sensitivity to brain activation that is extended over large regions, fine-grained patterns that may be present within the region of interest (ROI) are lost (Kriegeskorte et al., 2008; Kriegeskorte, Mur, & Bandettini, 2008; Albers et al., 2018).

In most univariate analyses, the blood oxygenation level dependent (BOLD) signal in each voxel (i.e. a three-dimensional volume of the brain, like a 3D pixel) is separately tested by comparing the values between conditions at the level of the individual subject (first-level analysis), typically by performing a t-test examining the null hypothesis that the expected response is not different between two conditions that are often each compared to a baseline or rest condition (Todd et al., 2013). Then, at the group (second-level) analysis, a random effects analysis is used to determine whether or not the average difference between two conditions is consistent across multiple subjects. For a particular voxel, if the mean difference between conditions is significantly different from zero, that voxel is said to be responsive at the group level. One important assumption for this type of analysis is that for a given experimental task, there is a similar spatial pattern of neural activity that is shared across participants such that the second level analysis is significant (Gilron et al., 2017). If there are significant individual differences in the spatial pattern of activity, even in the presence of strong individual effect sizes, there can be no significant effect at the group level. Given the variable and context-dependent nature of MSI and sound-symbolic CCs, to the extent that variation in spatial patterns of activation may also reflect variation in MSI, this assumption can be a problem for fMRI studies that explore these topics.

1.6.2 Multivariate analysis approaches

In contrast, multivariate analyses allow scientists to compare this spatial pattern of activity across individuals by examining and testing multiple voxels instead of each voxel individually (Haxby et al., 2001). Often, a machine learning approach like linear discriminant analysis or support vector machine learning is used, and the results compare a statistic, usually the measure of classification accuracy, against a null distribution that is centered around the level of chance classification (Kragel et al., 2012; Misaki et al., 2010; Mur, Bandettini, & Kriegeskorte, 2009; Tong & Pratte, 2012). Once this statistic is calculated at the individual subject level, it’s carried over to the second level. Different from the t-statistic or beta contrast from univariate analyses, the classification accuracy or other statistic in a multivariate analysis can be significant in the presence of individual variation of the direction of the effect or of the spatial distribution of the effect. The motivation to develop multivariate approaches to neural analysis methods was to reveal weak or distributed signals and information at finer spatial scales than are capable with univariate fMRI (Haxby, 2012; Haynes & Rees, 2006; Kamitani & Tong, 2005).

Multivariate analyses allow for multiple comparisons in large datasets without requiring the same assumptions that constrain typical GLM-based approaches looking at ROIs that have been defined by the experiment (Kriegeskorte & Keivit, 2013). One type of multivariate analysis, representational similarity analysis, or RSA, is gaining relevance in neuroscience due to reduction in assumptions compared to other univariate approaches and due to its flexibility as a method for identifying small scale patterns over space and time (Kriegeskorte, Goebel, & Bandettini, 2006). RSA has been used to investigate electroencephalography (EEG), magnetoencephalography (MEG), fMRI, and electrophysiology datasets for a multitude of comparisons (Hanke et al., 2009; Kriegeskorte et al., 2008; Tyler et al., 2013). The basic principle of RSA relies on the construction of a representational dissimilarity matrix (RDM) that compares all conditions being tested against one another and computes a dissimilarity value (1-r) and a significance value for each possible correlation (Kriegeskorte, 2009). RSA can be used to explore neural patterns among different ROIs (See Chapter 3) or it can be used to explore behavioral ratings or other physical measurements of a set of stimuli (See Chapter 2). The immense flexibility of RSA is one of its greatest advantages in that we can compare responses to stimuli across sensory modalities or across different physical measurements of these stimuli.

Another type of multivariate method is a multivoxel searchlight analysis, one type of multivoxel pattern analysis (MVPA), that is particularly useful for determining regions important to the task of interest and uses a classifier of multiple conditions within a certain voxel radius across the entire cortex (Kriegeskorte, 2009). For instance, Kriegeskorte et al. (2008) used MVPA and RSA to compare fMRI activity patterns in humans viewing a corpus of images and could compare the beta value obtained from the general linear model (GLM) for each image against each other image for a specific ROI, the inferior temporal cortex (IT). In this manner, the images could be categorized not based on an experimenter's subjective interpretation of each image compared to the others but instead were grouped based on the patterns of activation that they elicited compared to baseline levels. Kriegeskorte et al. (2008) further illustrate the use of RSA by performing a similar analysis using electrophysiology data recorded from macaque ITs. RSA was used to directly compare neuronal activity of humans and monkeys for each of the stimulus images. RSA has also been employed to compare different time points within a trial to decode what information may be encoded in specific brain regions over time (Tyler et al., 2013). Both RSA and searchlight MVPA are multivariate analyses that have many diverse uses in the field of neuroscience, particularly in examining effects that may be small and spatially or directionally variable across participants (Devereux et al., 2013). We use a combination of multivariate and univariate approaches to objectively define regions that exhibit MSI and regions that are important for the processing of sound-symbolic CCs. Using a combination of these analytical approaches allows us to determine spatial patterns of activity that are either shared (univariate) or potentially unique (multivariate) across our participants.

1.7 Dissertation Project

The proceeding Chapters 2,3, and 4 report a series of studies that attempt to expand on previous work on sound-symbolic CCs and MSI in three ways. First in Chapter 2, we investigate the physical parameters, including both visual parameters in shapes and acoustic parameters in pseudowords, which enhance or decrease the perception of sound-symbolic associations (List et al., BioRxiv; McCormick et al., 2015; Nielsen & Rendall, 2011). Second in Chapter 3, we investigate the neural correlates of sound-symbolic CCs (See Chapter 3; McCormick et al., 2018; Revill et al., 2014). Third in Chapter 4, we investigate the behavioral and neural evidence of variation in the perception of sound-symbolic CCs in a group of typical individuals and a grapheme-color synesthete (Chen et al., 2016; Lacey et al., 2016). If the variation in behavioral and neural responses to sound-symbolic associations is correlated across individuals, these results would provide evidence for a relationship between variation in our neural and behavioral measures that could reflect the context-dependent nature of MSI, at least for sound-symbolic CCs. By identifying the meaningful associations evoked by acoustic parameters of speech and visual parameters of shape, by investigating the cortical locations and behavioral significance of such mappings, and by describing the extent of variation in the individuals who make these associations, the current studies are designed to examine one aspect of the sound-to-meaning mappings in natural language.

In Chapter 2 (List et al., BioRxiv), in order to determine how the structure of speech corresponds with particular meanings, an analysis of which acoustic parameters within pseudowords are associated with certain basic visual sensory-perceptual parameters was completed using representational similarity analysis (RSA). Possible correlations between certain visual parameters or certain acoustic parameters and judgments of word meaning, on a spectrum from rounded to pointed, were analyzed.

Using information about the correspondences between certain speech sounds and visual shapes, in Chapter 3, an fMRI task was developed to determine if these associations have functional significance for language processing and MSI. Language users were asked to respond to auditory-visual stimulus pairings in which visual objects (rounded or pointed) were presented with pseudowords that had sound structures which either matched (congruent) or mismatched (incongruent) visual object shape. This task was designed to determine the neural underpinnings of how these sound-to-meaning mappings influence visual shape processing as suggested by previous work (Parise & Spence, 2008; Lupyan & Spivey, 2010; Lupyan & Ward, 2013; Revill et al., 2014; McCormick et al., 2018).

If auditory-visual multisensory integration underlies sensitivity to sound-symbolic crossmodal correspondences, then the sound structure of the pseudowords should influence how quickly participants can decide if the visual shape matches or is mismatched (Parise & Spence, 2009; McCormick et al., 2018). When the sound-symbolic properties of the pseudowords match the visual shape properties, then identification should be facilitated. When pseudowords mismatch visual object shape, interference between the sound–to-meaning associations and the visual sensory input is expected across participants. However, based on previous work and the basic principle of MSI that it is context-dependent, these sound-symbolic CCs may be dependent on the individual and their life experiences (Chen et al., 2016). Moreover, grapheme-color synesthesia is a strong example of variation in perception, and the presence of grapheme-color synesthesia has been shown to correlate with variation in the perception of sound-symbolic CCs (Lacey et al., 2016). Therefore, in Chapter 4, we examine the extent to which individual variation contributes to the perception of sound-symbolic CCs neurally and behaviorally in a group of typical individuals and in a grapheme-color synesthete.

Chapter 2

Acoustic and Visual Stimulus Parameters Driving Sound-symbolic Crossmodal Correspondences[1]

2.1 Abstract

It is often assumed that a fundamental property of language is the arbitrariness of the relationship between sound and meaning. Sound symbolism, which refers to non-arbitrary mapping between the sound of a word and its meaning, contradicts this assumption. Sensitivity to sound symbolism has been studied through crossmodal correspondences (CCs) between auditory pseudowords (e.g. ‘loh-moh’) and visual shapes (e.g. a blob). We used representational similarity analysis to examine the relationships between physical stimulus parameters and perceptual ratings that varied on dimensions of roundedness and pointedness, for a range of auditory pseudowords and visual shapes. We found that perceptual ratings of these stimuli relate to certain physical features of both the visual and auditory domains. Representational dissimilarity matrices (RDMs) of parameters that capture the spatial profile of the visual shapes, such as the simple matching coefficient and Jaccard distance, were significantly correlated with those of the visual ratings. RDMs of certain acoustic parameters of the pseudowords, such as the temporal fast Fourier transform (FFT) and spectral tilt, which reflect spectral composition, as well as shimmer and speech envelope that reflect aspects of amplitude variation over time, were significantly correlated with those of the auditory perceptual ratings. RDMs of the temporal FFT (acoustic) and the simple matching coefficient (visual) were significantly correlated. These findings suggest that sound-symbolic CCs are related to basic properties of auditory and visual stimuli, and thus provide insights into the fundamental nature of sound symbolism and how this might evoke specific impressions of physical meaning in natural language.

---

[1] Author’s note: The text for this chapter has been modified from a manuscript based on this work that was recently submitted for peer review, *Acoustic and visual stimulus parameters underlying sound symbolic crossmodal correspondences* (List et al., BioRxiv).

2.2 Introduction

It is commonly held that a fundamental property of language is that it is arbitrary, i.e. the sounds used to describe a referent do not necessarily resemble that referent (de Saussure, 2011, but see Joseph, 2015). Whether language is fundamentally arbitrary or whether “names ought to be given according to a natural process” has been debated since Plato’s *Cratylus* (Ademollo, 2011, p.32). Language does have significant aspects that are non-arbitrary, for example, iconicity, which refers to the resemblance between properties of a linguistic form and the sensorimotor and/or affective properties of referents (Perniss & Vigliocco, 2014). One type of iconicity is sound symbolism, which includes a broad set of phenomena that show a clear resemblance between properties of speech sounds and properties of their referents, such as onomatopoeia or Japanese mimetics (e.g. ‘gorogoro’ meaning the sound of a heavy object rolling around) (Imai & Kita, 2014).

Because sound symbolism embraces a broad set of non-arbitrary phenomena, it has proven difficult to study systematically, but sound-symbolic crossmodal correspondences (CCs) have provided researchers with a unique framework with which to study sound symbolism. A well-known example of sound-symbolic CCs was first described by Köhler (1929) in which individuals displayed a correspondence between a curvy, cloud-like shape and the pseudoword “baluma” and between an angular star-like shape and the pseudoword “takete”.

Sound-symbolic CCs constitute one type of CC, i.e., near-universally experienced associations between seemingly arbitrary stimulus features in different senses (Spence, 2011). CCs often occur between stimulus properties that are correlated (i.e. small animals often emit higher-pitched noises than large animals), which can lead to more efficient processing of sensory information. For example, large and small size in the visual domain are consistently associated with low- and high- pitched sounds (Gallace & Spence, 2006), and high and low elevation of visual stimuli are associated with high and low pitch in the auditory domain (Ben-Artzi & Marks, 1995).

Sound symbolism has been demonstrated across different languages (Blasi et al., 2016), cultures (Chen et al., 2016; Kantartzis et al., 2011), and even with children of pre-reading age (Imai et al., 2015; Maurer et al, 2006; Ozturk et al., 2013; Tzeng, Nygaard, & Namy, 2017). These studies show that the existence of sound-symbolic CCs in language is both prolific and robust. Further, language users are sensitive to sound-symbolic correspondences in that they actively utilize these associations to correctly assign meaning to foreign synonym-antonym pairs at above-chance levels (Nygaard, et al., 2009; Revill et al., 2014; Tzeng, Nygaard, & Namy, 2016). Sound-symbolic CCs also seem to play a role in language processing and early word learning (Imai & Kita, 2014). For example, individuals exhibit sensitivity to sound-symbolic crossmodal associations as early as four months of age (Ozturk et al., 2013), and recent studies have suggested that sound symbolism is important for specific word-to-meaning associations in young children with limited vocabularies (Gasser, 2004; Tzeng, Nygaard, & Namy, 2017). In adults, sound symbolism may offer linguistic processing advantages for categorization and word learning, particularly for larger vocabularies (Brand et al., 2018; Gasser, 2004; Revill et al., 2018). Though most of the evidence supporting sound symbolism has consisted of behavioral studies, more recently neuroimaging studies have begun to reveal the neural correlates of this phenomenon (Revill et al., 2014; McCormick et al., 2018; Peiffer-Smadja & Cohen, 2019; See Chapter 3).

Two open questions on sound-symbolic CCs are which physical features in the auditory domain come to be associated with which features in the visual domain; and how these features combine to drive a sound-symbolic association. Previous research has suggested that certain phonetic features are more likely to elicit sound-symbolic associations (Nielsen & Rendall, 2011; Fort, Martin, & Peperkamp, 2014; McCormick et al., 2015). For example, obstruents, including the English consonant sounds ‘p’, ‘t’, and ‘k’, are more likely to evoke pointedness, and sonorants, including the English consonant sounds ‘l’ and ‘m’, are more likely to evoke roundedness (McCormick et al., 2015). In addition, the relevant phonetic features can change depending on the visual association: Knoeferle et al. (2017) found that the first formant predicted visual size judgments, the third formant predicted visual shape judgments, and both size and shape judgments were predicted by the second formant. However, the acoustic and visual drivers of sound symbolism (e.g. the spectrotemporal features and amplitude variations of speech sounds and the spatial features of shapes) have received limited examination. In tasks that test artificial language learning, non-arbitrary word-meaning associations are easier to learn than arbitrarily constructed word-meaning associations (Brand et al., 2018; Nielsen, 2016; Revill et al., 2018), but researchers do not know what stimulus features participants attend to during the perception of these artificial words. Assessing the physical features that contribute to a sound-symbolic CC, as we do here, provides insight into the fundamental perceptual processes that drive word-to-meaning associations.

In the experiments reported here, we used sets of systematically varied auditory pseudowords and visual shapes, each of which was rated by participants along dimensions of roundedness and pointedness. Similarities among these perceptual judgments were compared between modalities, and to similarities among acoustic and visual parameters of the stimuli. In order to compare across auditory and visual modalities and between perceptual ratings and various parameters of the stimuli, we used representational similarity analysis (RSA; Kriegeskorte et al., 2008). This method relies on generating representational dissimilarity matrices (RDMs) and has been applied to capture the neural dissimilarity between individual stimuli or classes of stimuli (Kriegeskorte et al., 2008). Here we use RSA in a novel context by comparing the patterns of dissimilarity between judgments of systematically varied shape and sound stimuli and measurements of acoustic and visual parameters of these stimuli. RSA allows us to develop a comprehensive view of the commonalities and the differences in the dissimilarity patterns that relate to both visual and auditory stimuli in the perception of sound-symbolic CCs. The present study was designed to systematically determine which acoustic features best characterize sound-symbolic stimuli and which visual features are related to judgments of shape, using RSA. Specifically, we examined whether a key set of acoustic and visual parameters reliably predicted judgments of roundedness and pointedness of the sounds of pseudowords and the shapes of two-dimensional objects.

2.3 Materials and Methods

2.3.1 Participants

A total of 61 Emory University students (28 male, 33 female, mean (± standard deviation (SD)) age 20 ± 4 years) gave informed consent and received course credit for their participation. Thirty participated in a Likert-type rating task for visual shapes (14 male, 16 female) and a separate 31 participated in a Likert-type rating task for auditory pseudowords (14 male, 17 female). All participants were native English (American) speakers and reported normal or corrected-to-normal vision and no known hearing, speech, or language disorders. All procedures were approved by the Emory University Institutional Review Board.

2.3.2 Stimuli

2.3.2.1 Shapes

Ninety shapes were created using Adobe Illustrator (Ventura, CA) and consisted of gray line drawings (RGB: 240, 240, 240) on a black background. Shapes had four, five, or six protuberances. Shapes were created using a template of three concentric circles (25, 35, and 45 mm radii), the outer circle serving as a bounding border with protuberances extending to its perimeter. The two inner circles served to define the inward extent of each protuberance. Thinner protuberances (30 shapes) extended all the way to the innermost circle; thicker protuberances (30 shapes) extended only to the middle circle; the remaining 30 shapes were constructed with a mix between thin and thick protuberances. For each shape of one category (rounded or pointed), there was a corresponding shape in the other category with the same outer and inner anchor points, which created fifteen thick, fifteen thin, and fifteen heterogeneous shapes in each category. Under these constraints, the full complement of shapes included a rounded and pointed shape with the same inner and outer anchor points.

2.3.2.2 Pseudowords

570 pseudowords were constructed using only phonemes and combinations of phonemes that occur in the English language (McCormick et al., 2015). Each pseudoword consisted of two syllables of the form: consonant, vowel, consonant, vowel (CVCV). Consonants were sampled from sonorants, fricatives/affricates, and stops; of the obstruents, including fricatives/affricatives

and stops, half were voiced and half were unvoiced. Vowels were either front/rounded or back/unrounded. For a complete description of the stimulus set, see McCormick et al. (2015). The pseudowords were recorded by a female native speaker of American English with neutral intonation. Pseudowords were read from a computer screen and recorded in a sound-attenuated room using a Zoom 2 Cardioid microphone and digitized directly onto a Dell computer operating Windows Vista with an Intel Core2 processor at a 44.1 kHz sampling rate. Each pseudoword was then down-sampled at 22.05 kHz (standard for speech and to allow for consistent sample sizes in acoustic measurements of the pseudowords) and amplitude-normalized using PRAAT (Boersma & Weenink, 2012), a speech analysis software package. The pseudowords had a mean (± SD) duration of 457 ± 62 ms.

Each pseudoword was checked by four raters for accurate pronunciation of the intended segments and mispronounced pseudowords were re-recorded and edited, after which they were re-checked by the same four raters. Thirty-three of the pseudowords were rated by at least one of four independent listeners to be real English words and were therefore excluded, for a final total of 537 pseudowords (McCormick et al., 2015).

2.3.3 Likert-type rating tasks

The 537 pseudowords were presented auditorily, once each, over Beyerdynamic DT100 headphones at approximately 75db SPL for one set of participants (n = 31). For a separate set of participants (n = 30), the 90 shapes were presented sequentially, once each, at the center of a Windows 7.0 desktop computer screen using EPrime software Version 2.0.8.22 (Schneider, Eschmann, & Zuccolotto, 2002). The stimuli were presented in pseudorandom order.

Participants were randomly assigned to one of four Likert-type ratings tasks in which they were asked to rate pseudowords or shapes using one of two 7-point Likert-type scales: the first rated pointedness from 1 (not pointed) to 7 (very pointed) and the second rated roundedness from 1 (not rounded) to 7 (very rounded). For pseudowords, 16 participants used the pointedness scale and 15 the roundedness scale (total n = 31). To discourage participants from matching pseudowords with a specific word in the instructions (e.g. 'peh-kee' and 'pointed'), the instructions included several related terms for the concepts of rounded and pointed. For the shapes, 13 participants used the pointedness scale and 17 the roundedness scale (total n = 30). For both pseudowords and shapes, the 7-point scale appeared on the screen on each trial, either in the center of the screen for pseudowords or below each shape. The response keyboard always had 1-7 listed from left to right.

2.3.4 Data processing and analysis

2.3.4.1 Behavioral Analysis

Behavioral data were analyzed in MATLAB 2016a (The MathWorks, Natick, 2016), visual parameters were analyzed using MATLAB, and acoustic parameters were analyzed using both MATLAB and PRAAT (Boersma & Weenink, 2012). We implemented RSA using custom-built code in MATLAB to analyze how each of our acoustic and visual parameters of interest related to the perceptual RDMs. This method involves computing second-order correlations, between a reference RDM and test RDMs (Figure 2.1).

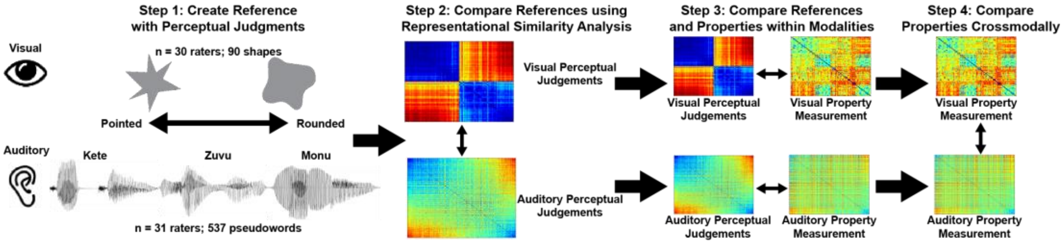


Figure 2.1. Analysis pipeline for Chapter 2. In step 1, we created reference RDMs based on the perceptual judgments in the Likert task. In step 2, we compared the perceptual judgments of the shapes to those of the pseudowords. In step 3, we then compared RDMs of perceptual judgments to RDMs of various measurements of the stimuli for both shapes and pseudowords. In step 4, we compared RDMs of selected visual parameters of the shapes to those of selected auditory parameters of the pseudowords.

To first create a reference RDM based on the perceptual judgments (Figure 2.1, step 1), the pseudowords were ordered from most pointed to most rounded based on the mean Likert rating of each stimulus. When ordering the stimuli for the perceptual RDM models, the two Likert scales (1-7: not pointed to very pointed, and 1-7: not rounded to very rounded) were recoded such that 1 was equal to “not pointed” on one scale and “very rounded” on the other, and 7 was equal to “very pointed” on the first scale and “not rounded” on the second. However, once the stimuli were thus ordered, we used the original (un-recoded) data within the perceptual RDMs because the RDMs rely on variation, and we are here interested in comparing across pseudowords or across shapes, rather than comparing across participants: regardless of the scale that any individual participant used to rate the stimuli, we are interested in how the patterns of the judgments were dissimilar across the pseudowords or shapes. As illustrated by Figure 2.2, the judgments for the pseudowords and for the shapes were comparable across the two Likert scales.

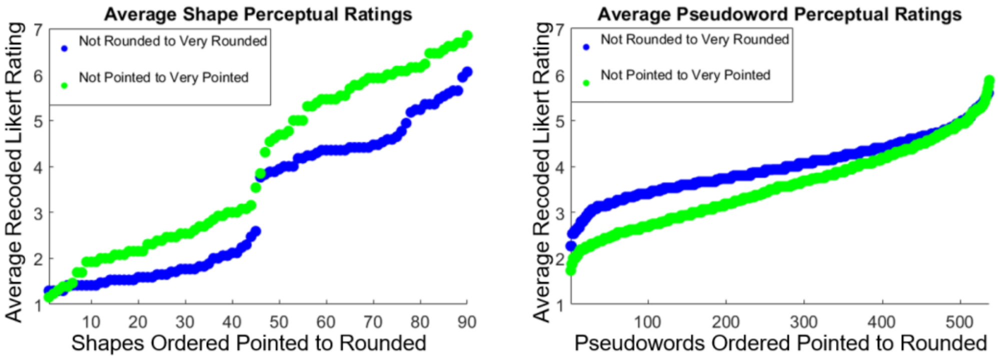


Figure 2.2. The recoded average rating for each shape (left) and pseudoword (right) stimulus across the two Likert scales.

First, we compared the perceptual judgments RDMs across modalities (Figure 2.1, step 2). The RDMs represent the pairwise dissimilarity between stimuli, given by 1- Pearson's correlation coefficient. Then we compared auditory and visual perceptual RDMs to their respective within-modal property RDMs (Figure 2.1, step 3). Using all participant ratings for shapes and for pseudowords respectively, we constructed two RDMs: one of 90 by 90 for shapes and the other of 537 by 537 for pseudowords. Each of these RDMs was used to compare with RDMs for specific parameters of the corresponding modality (see below).

For the pseudowords, when comparing perceptual judgments to parameters that comprised a single value per pseudoword using second-order correlations (see Table 2.4), we down-sampled the RDM for the perceptual judgments and binned the acoustic measurements to create 18x18 matrices with 30 pseudowords within each cell. This provided a sufficient number of measures to compute a correlation coefficient that could be entered into each cell of an RDM, while still allowing a sizable enough matrix to be compared with another RDM.

Finally, we compared significant shape and pseudoword parameters across modality (Figure 2.1, step 4). To compare across modalities, we down-sampled the RDM for pseudowords, which was already ordered via the perceptual ratings, by selecting every sixth pseudoword from the original 537 by 537 matrix to create a 90 by 90 matrix for pseudowords that could then be compared with the 90 by 90 matrix for shapes. To compare RDMs, we used Spearman's r for second-order correlations instead of Pearson correlations because Spearman's r is a non-parametric measurement that does not rely on the assumption of normality of the data.

2.3.4.2 Analysis of stimulus parameters

After analyzing the behavioral results and creating the perceptual RDMs, several visual and acoustic parameters of the stimuli were measured and used to generate RDMs for each parameter (Figure 2.1, step 3). To test the relationship of the different stimulus parameters to the perceptual ratings in each modality, second-order Spearman correlations between the RDMs for each parameter and the perceptual ratings in that modality were computed. Note that not all stimulus parameters required the same number of samples per image or per pseudoword, respectively, which may serve important when comparing effect sizes in the results (Tables 2.1 and 2.2).

We also compared selected stimulus parameters across modalities (Figure 2.1, step 4) in order to determine which were most likely to be driving sound-symbolic CCs. Once we determined which stimulus parameters had RDMs that were most correlated with those of the perceptual judgments, we used the RDMs of the two most highly correlated parameters from the visual domain to compare with highly correlated parameters from the auditory domain. We chose the two most highly correlated parameters in the auditory domain that would still allow us to maintain a consistent matrix size of 90 by 90 across modalities. We then compared RDMs of the two stimulus parameters in each modality with the two in the other modality.

2.3.4.2.1 Shapes: The visual parameters examined for the shapes were the Jaccard distance, the simple matching coefficient, and the fast Fourier transform (FFT) in the spatial frequency domain, all of which index spatial differences among the shapes, together with the silhouettes, binary images and image outlines, all of which attempt to account for shape in a pixel-wise manner (see Table 2.1 for computational details).

The Jaccard distance and simple matching coefficient (the latter is also known as the Rand similarity coefficient) are two types of distance measures that have been used to compare similarity and diversity, including patterns in DNA (Deagle et al., 2017), images (Devereux et al., 2013), and voices (Cernak et al., 2016). The spatial frequency FFT has been used to filter noisy or complex patterns in complex image sets (Petrou & Petrou, 2011). The spatial FFT has also been used to access the geometric characteristics of complex images, such as determining the spatial characteristics of retinal microglia (Zhang et al., 2016). The silhouettes of images have been shown to predict activation patterns in early visual cortex (Kriegeskorte et al., 2008) and have been used to reveal overlap in the neural processing of words with the processing of objects (Devereux et al., 2013). Binary images and image outlines capture the entire shape in the form of a vector but may be prone to noisiness that does not capture the spatial patterns among different shapes (Devereux et al., 2013). For all visual parameters, shapes were first rotated such that they aligned as closely as possible by co-registering the points or the center of the curves of each shape with one another.

Visual Parameters	**Description**
Jaccard Distance	The Jaccard distance has been used for RSA of images (Devereux et al., 2013) and is defined as one minus the size of the intersection of the shapes divided by the size of the union of two sample shapes. The Jaccard distance, a variation on the simple matching coefficient, more specifically is 1 - (a)/(a+b+c), where the components include the number of pixels present (a) in both images, (b) in the first image but not in the second, (c) in the second image but not in the first (Ricotta & Pavoine, 2015). This pairwise measure is included in the RDM for this parameter as the Jaccard distance instead of 1- Pearson r (160,000 samples/image).
Simple Matching Coefficient	The simple matching coefficient is an example of the use of presence/absence coefficients based on the matching and mismatching pixels within two images. The components of these coefficients include the number of pixels present (a) in both images, (b) in the first image but not in the second, (c) in the second image but not in the first, and (d) the pixels absent from both images but found in other images, such that the sum S = a + b + c + d is the total number of pixels in the entire collection of stimuli (Ricotta & Pavoine, 2015). The simple matching coefficient is given by (a+d)/(a+b+c+d). This pairwise measure is included in the RDM for this parameter as 1 – simple matching coefficient instead of 1- Pearson r (160,000 samples/image).
Spatial Frequency FFT	The spatial frequency FFT captures how often the components from the FFT repeat within an image per unit distance (cycles per millimeter; 40,000 samples/image).
Silhouette	Each of the shapes was converted into a series of zeros (white) and ones (black) such that each grey shape was converted to white so that it appeared as a silhouette against a black background. The shapes were then reduced from a two-dimensional shape into a single one-dimensional vector, and the RDM was constructed as for image outlines (see below; 40,000 samples/image).
Binary Images	Each of the shapes was converted into a series of zeros (white) and ones (black) such that both the outline (white) and silhouette (black) were included. The shapes were then reduced from a two-dimensional shape into a single one-dimensional vector, and the RDM was constructed as for image outlines (see below; 160,000 samples/image).

Image outlines	Each of the shapes was converted into a series of zeros (white) and ones (black) in which only the outline of each shape was included. The shapes were then reduced from a two-dimensional shape into a single one-dimensional vector. The resulting vectors were used to compute 1 – Pearson r for the RDM (160,000 samples/image).

Table 2.1. Description of visual parameters of shapes that were compared to the reference RDM from the perceptual judgments.

2.3.4.2.2 Pseudowords: The auditory parameters examined for the pseudowords (see Table 2.2 for computational details) were the speech envelope, spectral tilt, and FFT (here in the temporal frequency domain) all of which capture the complexity of the sound structure of each pseudoword across multiple measures of time or of frequency. By contrast, jitter, shimmer, pitch standard deviation, mean HNR, the fraction of unvoiced frames, mean autocorrelation, pulse number, and period number are all measurements that capture a single value for each pseudoword that represents an average or overall measurement of different parameters of the speech signal. For example, jitter and shimmer capture average variations in the glottal waveform, the standard deviation of the pitch captures average variation in the fundamental frequency, and mean HNR captures the average relative periodicity of the speech waveform for a given pseudoword. Even these parameters that arguably capture less complexity because they only provide a single measurement could give insight into the parameters that drive sound-symbolic judgments in the auditory domain. For example, the fraction of unvoiced frames is a measure of voicing, which was found to be important in our phonetic analysis of these pseudowords such that unvoiced pseudowords were more likely to be perceived as pointed (McCormick et al., 2015).

The speech envelope is a measure of the amplitude profile across time and captures energy changes corresponding to phonemic and syllabic transitions (Aiken & Picton, 2008). The speech envelope can visually capture the ‘shape’ of the pseudowords (see Figure 2.1, step 1, for examples), and is a clear candidate for driving sound-symbolic CCs. Spectral tilt contributes to the intelligibility of a speech signal under noisy conditions (Lu & Cooke, 2009). The temporal FFT reflects the power spectrum of the component frequencies and was measured at successive time points in the speech signal to capture the change in frequency composition across time (Singh, 2015). Jitter can be used to identify voice pathologies and stems from variation in the vibration of the vocal cords (Teixeira & Fernandes, 2014), giving rise to instability of sound frequency. Higher shimmer is correlated with the presence of noise and breathiness in the voice (Boersma & Weenink, 2012; Teixeira & Fernandes, 2014) and results from instability of sound amplitude. The standard deviation of the pitch or fundamental frequency has been implicated as an indicator of emotional and mental state, indicating that this measure can acoustically convey meanings that are not limited to the linguistic content (Kliper et al., 2015). The HNR gives an indication of the overall periodicity of a sound and has been used as an index of vocal aging or hoarseness (Ferrand, 2002). Mean autocorrelation is also a relative measure of periodicity of a signal. Since our stimuli were spoken by the same speaker taking care to keep prosody consistent, fundamental frequency should have remained relatively consistent, but pulse number may indicate variation in periodicity or phonation across stimuli (Hollien et al., 1977). All pseudowords were first normalized to a duration of 500 ms. Except where noted for the parameters of speech envelope, spectral tilt, and temporal FFT, RDMs comprised of a single sample or measurement per pseudoword.

Acoustic Parameters	**Description**
Speech Envelope	The speech envelope was calculated using the root mean square of the amplitude and a sliding window length of 120 samples (Hamming window of length 0.020 ms; 2250 samples/pseudoword).
Spectral Tilt	Spectral tilt is an estimate of the spectrogram (amplitude over multiple frequencies across time) and was calculated using the slope of the periodogram (power across frequency) at 225 equidistant points across each pseudoword, a window length of 600 samples (Hamming window of length 0.020 ms; 225 samples per pseudoword), and an overlap of 15 samples (length 0.020 ms).
Temporal frequency FFT	The FFT was calculated with a sliding time window of 600 samples (Hamming window of length 0.020 ms; 60,000 samples/pseudoword) and a frequency window of 100 Hz from 0 to 22 kHz.
Jitter	Jitter is the absolute difference between consecutive periods of the speech waveform divided by the average period of the speech waveform (Boersma & Weenink, 2012).
Shimmer	Shimmer is the average absolute difference between the amplitudes of consecutive periods of a sound wave, divided by the average amplitude, here measured in decibels (Boersma & Weenink, 2012).
Pitch Standard Deviation	The standard deviation of the pitch indicates the variation in the fundamental frequency present in the pseudowords.
Mean HNR	The HNR quantifies the ratio between the periodic, or harmonic, part and the aperiodic, or noisy, part of a voice signal and gives an indication of the overall periodicity of the sound (Teixeira & Fernandes, 2014).
Fraction of Unvoiced Frames	The fraction of unvoiced frames is the percentage of measurement windows (length 0.01 seconds) that do not engage the vocal folds (Boersma & Weenink, 2012).
Mean Autocorrelation	Mean autocorrelation is the average correlation of a sound with a delayed copy of itself and is a relative estimate of periodicity wherein 0 is a white noise signal and 1 is a perfectly periodic signal (Boersma & Weenink, 2012).

Pulse Number	The number of glottal pulses during production of vowels or voiced consonants (Boersma & Weenink, 2012).
Period Number	This measurement is the number of peak-to-peak cycles within the fundamental frequency of the sound signal used to calculate pitch (Boersma & Weenink, 2012).

Table 2.2. Description of acoustic parameters of the pseudowords that were compared to the reference RDM from the perceptual judgments.

2.4 Results

2.4.1 Comparison of visual and auditory perceptual ratings

First, we compared visual and auditory perceptual ratings across modalities (Figure 2.1, step 2). The pattern of judgments evident in the 90x90 RDM for the pointed/roundedness ratings for the visual shapes (Figure 2.3, left) was essentially binary. This pattern indicates that participants rated shapes largely as more pointed or more rounded, with only a few shapes considered intermediate. Figure 2.3 (right) depicts the sub-sampled 90x90 RDM of pointed/roundedness ratings for the pseudowords, which appeared to be more graded. Yet, the two matrices for the shapes and pseudowords were positively correlated (Spearman $r = 0.66$, $p<0.0001$). This finding indicates that, even though independent sets of participants judged shapes and pseudowords, the judgments were crossmodally consistent. One may expect this result, given that participants were asked to rate stimuli along dimensions of pointedness or roundedness in both modalities, but for the large set of pseudowords, consistent ratings using a scale that is primarily defined by vision was not guaranteed. The strong positive correlation between the auditory and visual RDMs establishes the presence of sound-symbolic CCs between a large set of 90 auditory pseudowords and 90 visual shapes. This lays the foundation for our subsequent analyses of the stimulus parameters relevant to the CC.

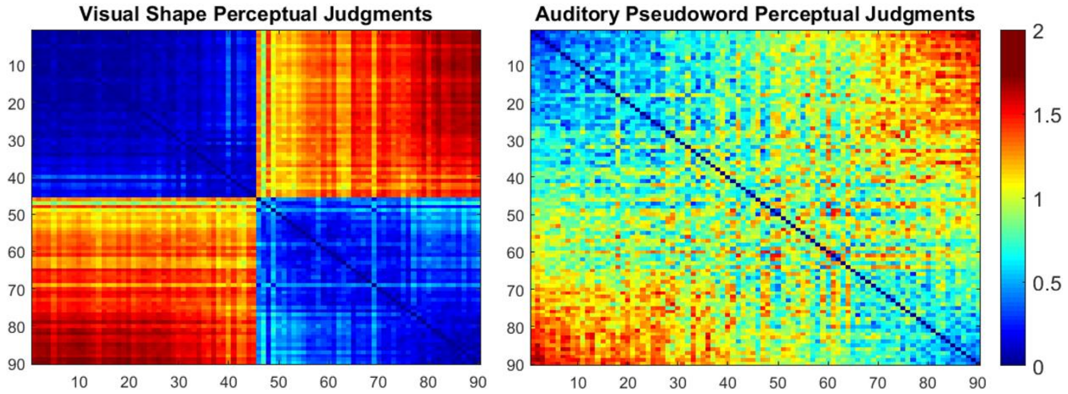


Figure 2.3. RDMs ordered according to mean perceptual judgments for visual shapes (left) and auditory pseudowords (right) from 1 (rated most pointed) - 90 (rated most rounded). Color bar shows dissimilarity increasing from 0 - 2, which applies to all figures with RDMs. Based on the pattern of correlations, visual shape judgments tended to be binary. Although some pseudowords were judged to be intermediate, there was a significant positive correlation between the two RDMs (Spearman r = 0.66, p < 0.0001).

2.4.2 Comparison of visual perceptual ratings to visual shape indices

Then we compared the visual perceptual RDM to its within-modal visual parameter RDMs (Figure 2.1, step 3). To determine what properties may influence the judgment of visual roundedness or pointedness, we next compared the RDM for visual judgments to each of the RDMs for the quantitative indices of shape. As an example, Figure 2.4 compares the dissimilarity matrix of pointedness/roundedness ratings for the shapes (left) to that for the simple matching coefficient (right). These two matrices were positively correlated (Spearman r = 0.28, p<0.0001), although the perceptual judgment RDM was clearly binary in comparison to that for the simple matching coefficient.

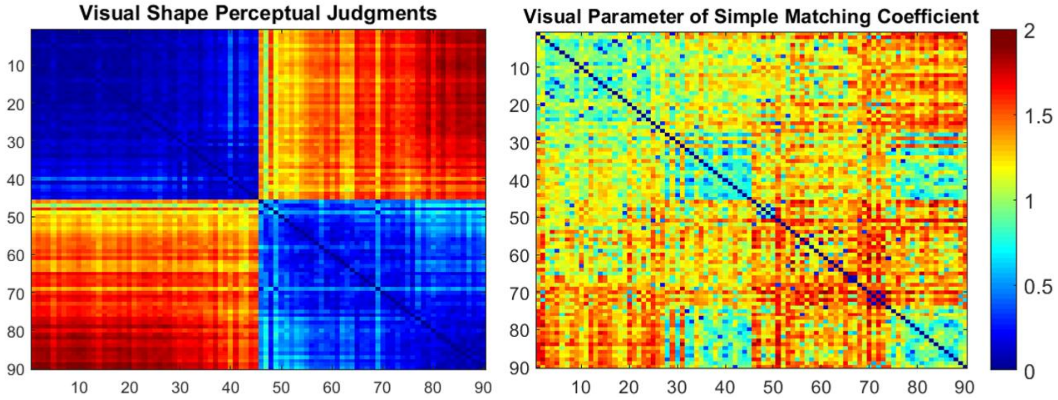


Figure 2.4. RDMs for the perceptual judgments for visual shapes (left) and one index of shape, the simple matching coefficient (right). Both RDMs are ordered according to mean perceptual judgments for shapes from 1 (rated most pointed) - 90 (rated most rounded). There was a positive correlation between the two RDMs, indicating that the simple matching coefficient is related to visual perceptual judgments of pointedness/roundedness (Spearman r = 0.28, $p < 0.0001$).

Table 2.3 shows that the RDM for shape judgments was positively correlated not only with that for the simple matching coefficient as noted above, but also with the RDMs for the silhouette, and Jaccard distance, indicating that these indices of the shapes represent parameters that are important for the visual judgment of roundedness or pointedness. Note that not all of the visual indices required the same number of samples per image, which may affect the comparison of effect sizes (Methods, Table 2.1). The RDMs for binary images, spatial FFT, and image outline were not significantly correlated with the RDM for shape judgments, indicating that these parameters are less important for the judgment of visual rounded or pointedness.

Visual Indices	**Second Order Spearman Correlation to Perceptual Ratings**
Simple Matching Coefficient	0.28 *
Silhouette	0.14 *
Jaccard Distance	0.10 *
Spatial FFT	0.02
Binary Images	0.06
Outline of Images	- 0.12

Table 2.3. Correlations between RDMs for visual perceptual judgments and quantitative parameters of shape * significant correlations $p < 0.0001$, Bonferroni-corrected α for six correlation tests = 0.0083.

2.4.3 Comparison of auditory perceptual ratings to acoustic parameters of pseudowords

Then we compared the auditory perceptual RDM to its within-modal acoustic parameter RDMs (Figure 2.1, step 3). To determine what auditory stimulus properties may influence the judgment of roundedness or pointedness of pseudowords, we compared the 537x537 RDM for pseudoword judgments with the RDMs for the quantitative indices of these pseudowords. For the following comparisons, we returned to the original set of 537 pseudowords to examine the entire spectrum of rounded/pointedness and the acoustic parameters that may influence these judgments. As an example, Figure 2.5 compares the dissimilarity matrix of pointedness/roundedness ratings for the pseudowords (left) to that for the temporal FFT: these matrices were positively correlated (Spearman r = 0.28, $p < 0.0001$).

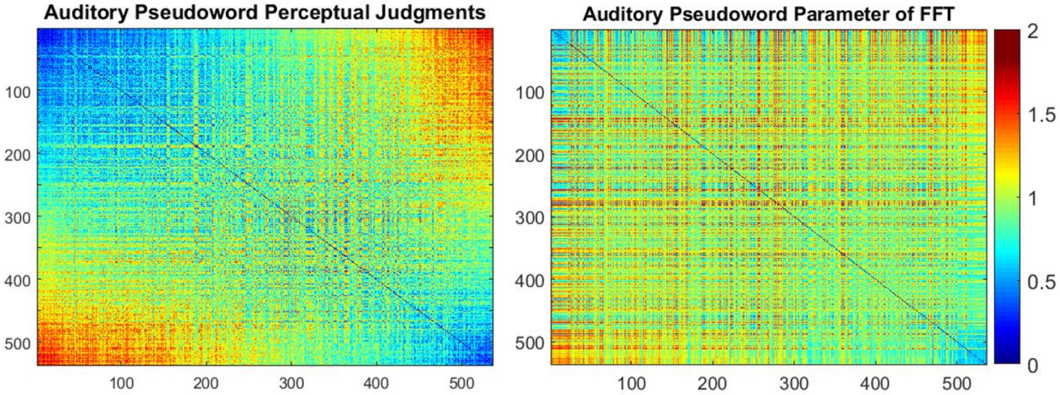


Figure 2.5. RDMs for the perceptual judgments for pseudowords (left) and one acoustic parameter, the temporal FFT (right). Both RDMs are ordered according to mean perceptual judgments for pseudowords from 1 (rated most pointed) - 537 (rated most rounded). A positive correlation was found between the two RDMs, indicating that the temporal FFT captures acoustic parameters important for auditory perceptual judgments of pointedness/roundedness (Spearman r = 0.28, p < 0.0001).

Table 2.4 shows that the RDM for pseudoword judgments was positively correlated with the RDMs for shimmer, the temporal FFT (as noted above), mean HNR, speech envelope, and spectral tilt, indicating that these acoustic parameters capture properties that are important for the judgment of roundedness or pointedness for pseudowords. The RDMs for the fraction of unvoiced frames, jitter, mean autocorrelation, period number, pulse number, and pitch standard deviation were not significantly correlated with those for judgments of pseudowords, indicating that these parameters are less important for the judgment of roundedness or pointedness for auditory pseudowords. When comparing perceptual judgments to parameters that comprised a single value per pseudoword using second-order correlations (Methods Table 2.2), we down-sampled the RDM for the perceptual judgments and binned the acoustic measurements to create 18x18 matrices with 30 pseudowords within each cell. This provided a sufficient number of measures to compute a correlation coefficient that could be entered into each cell of an RDM, while still allowing a sizable enough matrix to be compared with another RDM. Because of the requisite binning procedure for the most highly correlated parameter of shimmer, we instead chose the most highly correlated parameter that was not binned, the temporal FFT, as the example RDM in Figure 2.5 to illustrate the entire spectrum of pseudoword correlations.

Acoustic Parameters	**Second Order Correlation to Perceptual Ratings**
Shimmer	0.30 *
Temporal FFT	0.28 **
Mean HNR	0.27 *
Speech Envelope	0.14 **
Spectral Tilt	0.10 **
Fraction of Unvoiced Frames	0.10
Jitter	-0.01
Mean Autocorrelation	-0.03
Period Number	-0.04
Pulse Number	-0.08
Pitch Standard Deviation	-0.12

Table 2.4. Correlations between the RDMs for auditory perceptual judgments and quantitative parameters of pseudowords. Significant correlations = * $p < 0.001$; **$p < 0.0001$, Bonferroni-corrected α for eleven correlation tests =0.0045.

2.4.4 Comparison of indices of visual and acoustic stimuli

We next compared significant shape and pseudoword parameters across modality (Figure 2.1, step 4). Because the visual shape and auditory pseudoword judgments involved independent sets of participants, it is possible that participants were making judgments based only on a conceptual representation of rounded or pointed that did not involve sound-symbolic crossmodal associations. If participants were not employing sound-symbolic CCs in their judgments of the pseudowords, we would expect the visual and acoustic parameters that were significant for the two sets of judgments to be unrelated. If instead participants were using a common perceptual framework while judging stimuli from a single modality, particularly the auditory pseudowords, we would expect the auditory and visual parameters to be correlated with one another.

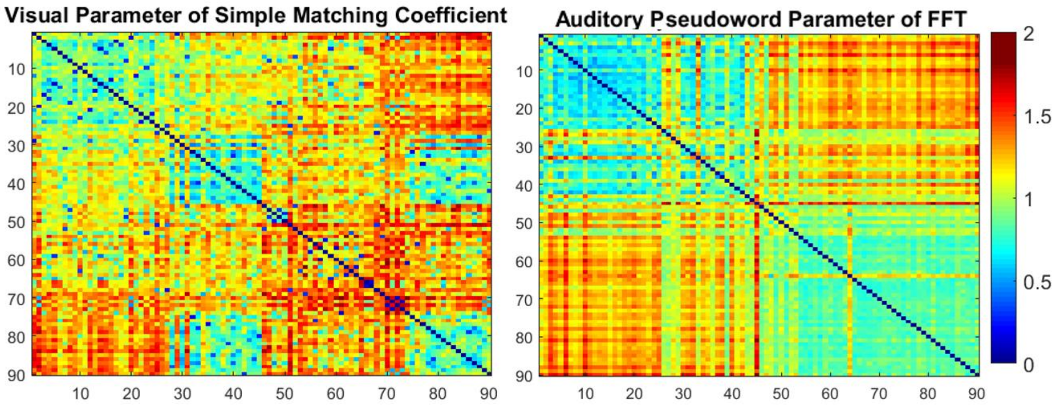


Figure 2.6. RDMs for the visual simple matching coefficient (left) and the acoustic temporal FFT (right). These RDMs are ordered according to mean perceptual judgments for shapes (left) or pseudowords (right) from 1 (rated most pointed) - 90 (rated most rounded). A positive correlation was found between these RDMs, indicating that the simple matching coefficient for the shapes and the FFT for the pseudowords both relate to properties relevant for the perception of sound-symbolic CCs ($r = 0.13$, $p < 0.0001$).

We compared stimulus parameters across modalities in order to determine which of the examined parameters in the two modalities are most likely to be driving sound-symbolic CCs (Figure 2.6 and Table 2.5). Once we determined which stimulus parameters were most correlated with the perceptual judgments, we used the RDMs of the two most positively correlated parameters from the visual domain to compare with the two most positively correlated parameters from the auditory domain that would still allow us to maintain a consistent matrix size of 90 by 90. Thus, we chose the simple matching coefficient and silhouette from the visual parameters, and the FFT and speech envelope from the acoustic parameters.

Table 2.5 shows that the RDM for the simple matching coefficient was significantly correlated with that for the temporal FFT (Spearman $r = 0.13$, $p<0.001$: Bonferroni-corrected α for four correlation tests = 0.0125), which suggests that the temporal FFT is an acoustic parameter capturing aspects of the signal that are related to the visual parameters of the shapes indexed by the simple matching coefficient. The silhouette was also weakly positively correlated with the FFT (Spearman $r = 0.03$, $p<0.05$) but this correlation did not survive Bonferroni correction. The RDMs for the simple matching coefficient and silhouette were not significantly correlated with that for the speech envelope (all $p > .2$), indicating that, while the speech envelope is important for auditory judgments of roundedness and pointedness, it may not be directly related to the visual parameters of the simple matching coefficient and silhouette.

Visual Parameters	Auditory Parameters	Second-Order Spearman Correlation Visual to Auditory Parameters
Simple Matching Coefficient	Temporal FFT	0.13*
Silhouette	Temporal FFT	0.03
Simple Matching Coefficient	Speech Envelope	- 0.01
Silhouette	Speech Envelope	0.02

Table 2.5. Correlations between visual and acoustic parameters. Significant correlations = * $p < 0.0001$, Bonferroni-corrected α for four correlation tests = 0.0125.

2.5 Correlations between significant acoustic parameters and auditory perceptual ratings

For the acoustic parameters that yielded RDMs that were correlated with the RDM for the perceptual judgments, we were interested in how these parameters relate to what makes a ‘round’ or a ‘pointed’ pseudoword. Our parameters of shimmer and mean HNR included one measurement per pseudoword, so we were able to take a quantitative approach in correlating shimmer or mean HNR, respectively, with the mean perceptual rating of each pseudoword.

As noted earlier, higher shimmer can indicate increased hoarseness or breathiness in the voice (Teixeira & Fernandes, 2014) and is a measure of amplitude instability in the glottal waveform. We found that shimmer decreased from more pointed to more rounded pseudowords ($r = -0.40$, $p<0.0001$) indicating that a pseudoword that is judged as rounded has lower shimmer than one that is judged as more pointed. Mean HNR is a measure of the relative amount of noisiness in the signal with higher values indicating less noise. Thus, an increase in HNR may indicate increased ‘smoothness’ of the waveform (Teixeira & Fernandes, 2014). In contrast to shimmer, there was a positive correlation between mean HNR and ratings from more pointed to more rounded pseudowords (r = 0.53, p<0.0001), indicating that pseudowords that were judged as more rounded had a higher mean HNR than those that were judged as more pointed.

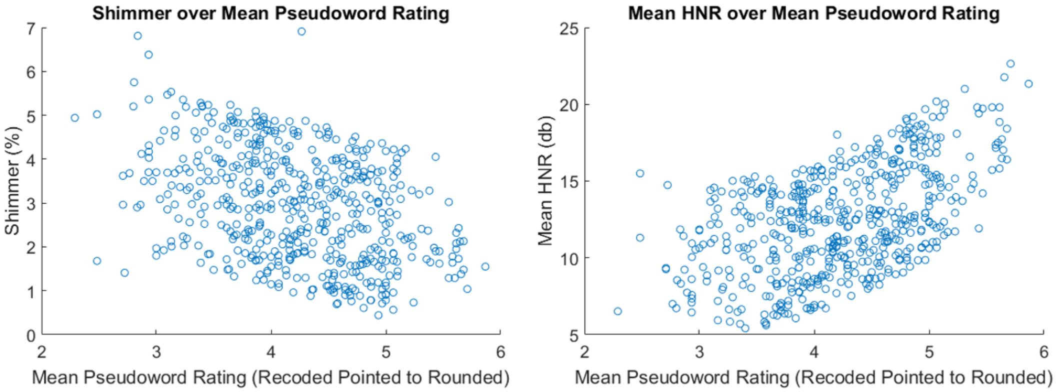


Figure 2.7. Scatterplots for the acoustic parameters of shimmer (left) and mean HNR (right) over the mean pseudoword ratings recoded across Likert tasks to range from 1-7: pointed to rounded (from left to right on the x axes). Shimmer was negatively correlated, and mean HNR was positively correlated with ratings.

Our parameters of temporal FFT, speech envelope and spectral tilt all included multiple measurements per pseudoword, preventing us from performing a simple correlation between each of these parameters and the pseudoword ratings. Instead, we took a qualitative approach in illustrating how these parameters might vary between a more pointed pseudoword (‘keh-teh’) versus a more rounded pseudoword (‘moh-loh’). For example, the temporal FFT is a recreation of the spectrogram for each pseudoword. Illustrated in Figure 2.8 (top) is the waveform for ‘keh-teh’ (left) and ‘moh-loh’ (right). The spectrogram (Figure 2.8, bottom) of the two pseudowords visually illustrates what is captured by the FFT, including the amplitude (degree of dark shading) across time (x-axis) over multiple frequencies (y-axis) for each pseudoword. Figure 2.8 shows the unvoiced (aperiodic sections of both the spectrogram and the waveform) portion of the pointed pseudoword ‘keh-teh’. Figure 2.9 illustrates the power spectral density (PSD) of the pointed pseudoword ‘keh-teh’ (left) versus the rounded pseudoword ‘moh-loh’ (right); the slope of the PSD over short time windows illustrates the spectral tilt, another complex acoustic parameter that was important for sound-symbolic judgments of pseudowords. Both Figure 2.8 and Figure 2.9 capture the increased power at higher frequencies in the pointed versus the rounded pseudowords.

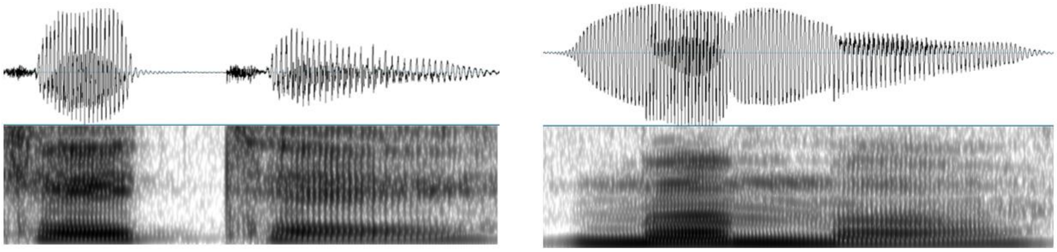


Figure 2.8. Waveform (top) and spectrogram to illustrate the acoustic parameter of temporal FFT (bottom) for a pointed pseudoword ‘keh-teh’ (left) and a rounded pseudoword ‘moh-loh’ (right). More abrupt changes in the power, especially at higher frequencies, for the pointed pseudoword ‘keh-teh’ versus the rounded pseudoword ‘moh-loh’ were captured by the temporal FFT.

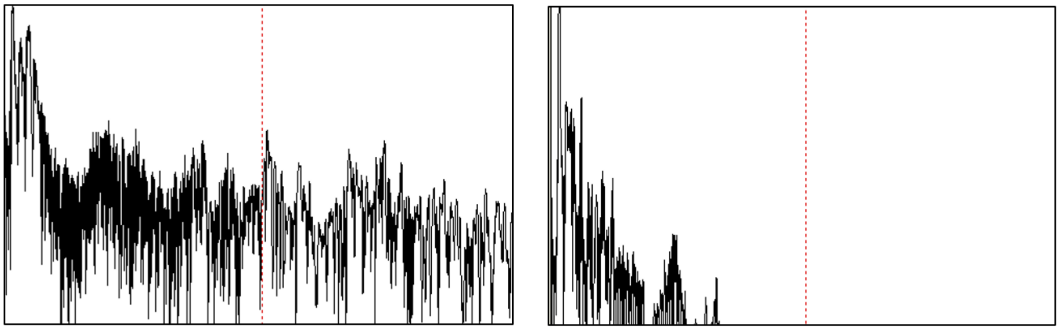


Figure 2.9. PSD showing power across time for two pseudowords illustrates the acoustic parameter of spectral tilt for a pointed pseudoword ‘keh-teh’ (left) and a rounded pseudoword ‘moh-loh’ (right). The PSD and spectral tilt capture increased power at higher frequencies for the pointed pseudoword ‘keh-teh’ versus the rounded pseudoword ‘moh-loh’.

Finally, we qualitatively examined the speech envelope for a pointed pseudoword (‘keh-teh’, Figure 2.10, left) versus a rounded pseudoword (‘moh-loh’, Figure 2.10, right). The speech envelope encapsulates the shape of the waveform in continuous measurements of amplitude across time. This measurement is perhaps the most intuitive illustration of the ‘shape’ of a pseudoword and how the jagged shape of a pointed pseudoword (left) differs from the rolling, full shape of a rounded pseudoword (right).

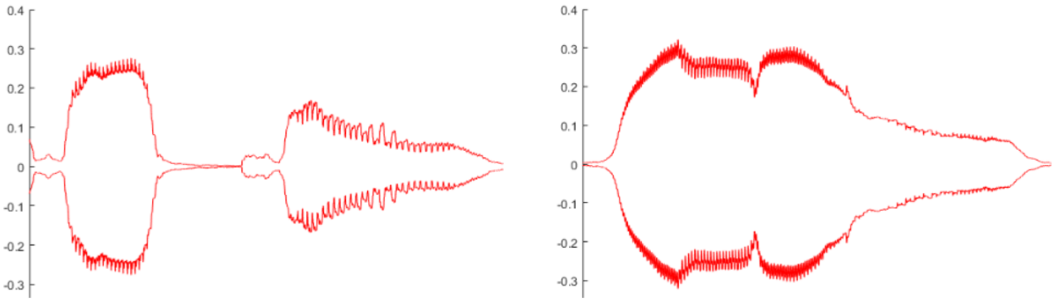


Figure 2.10. The acoustic parameter of speech envelope for a pointed pseudoword ‘keh-teh’ (left) and a rounded pseudoword ‘moh-loh’ (right) illustrates the “shape of sound” as measured by the amplitude across time.

2.5 Discussion

2.5.1 Sound-symbolic crossmodal correspondences and perception

The present study is the first to systematically examine both the acoustic and the visual drivers of sound symbolism in the sound-shape domain within the same experimental paradigm and using RSA; previous studies have only examined such parameters separately (visual: Chen et al., 2016; acoustic: Knoeferle et al., 2017). Additionally, our stimulus sets comprising 90 shapes and 537 pseudowords are more extensive than in previous studies (e.g. Knoeferle et al., 2017 included 10 shapes and 100 pseudowords). This advantage allowed us to explore a spectrum of rounded to pointed shapes and pseudowords, with significant power. Our study is also the first to investigate the similarities in perceptual judgments of rounded/pointedness for shapes compared to pseudowords and to further use these similarities to determine which visual and acoustic parameters are important for sound-symbolic CCs both within and across modalities. Despite using different parameters to make sound-symbolic judgments, raters were able to use a common scale of roundedness/pointedness across the two different modalities, as indicated by the similarity in the RDMs for shape versus pseudoword judgments. This is not necessarily surprising, given that participants were asked to use the same scales in both modalities, but the fact that different groups of participants, using two opposing scales in each modality, produced ratings whose dissimilarity matrices were highly correlated across modalities, is evidence that the sound-symbolic CC studied here is robust. The strong crossmodal correlation provides a firm basis for further exploration of the principal aim of the present study, characterizing stimulus parameters that may underpin sound symbolism.

For the shapes, the parameters of simple matching coefficient, silhouette, and Jaccard distance were predictive of individuals' judgments of visual roundedness/pointedness, as revealed by correlations between the RDMs for these parameters and the perceptual ratings. For the pseudowords, RDMs showed that the parameters of shimmer, temporal FFT, mean HNR, speech envelope, and spectral tilt were predictive of individuals' judgments of the roundedness/pointedness of sounds. As further illustrated by inspecting the spectrograms for the temporal FFT and spectral tilt for pseudowords judged as very pointed and very rounded, increased power in higher frequencies led to perceptual ratings that were more pointed. Inspecting the corresponding speech envelopes indicated that smoother and more gradual changes in amplitude were associated with perceptual ratings that were more rounded. Examination of correlations revealed that the acoustic parameter of shimmer, which can indicate increased breathiness (Teixeira & Fernandes, 2014) and amplitude variation and instability, *decreased* progressively as the ratings of roundedness increased, whereas the acoustic parameter of mean HNR, which indicates the ratio of periodic or smooth to aperiodic or noisy parts of a vocal signal, *increased* progressively as the ratings of roundedness increased.

Using RSA to compare across acoustic and visual parameters, we were able to determine that the spectral composition of auditory pseudowords as measured by the temporal FFT may be employed by listeners in a similar way as the visual index of shape quantified by the simple matching coefficient. This evidence suggests that even for unisensory stimuli, a parameter of sound in the case of pseudowords can be perceived as equivalent to a parameter of vision in the case of shapes, offering a basis for the sound-symbolic pseudoword-shape CC.

Our findings extend previous studies on sound symbolism suggesting that sound-symbolic ratings are grounded in broad categorical contrasts, such as the phonological features characterizing obstruent (e.g. ‘p’, ‘t’, and ‘k’ sounds in English) or sonorant (e.g. ‘m’, ‘n’, or ‘l’ sounds in English) consonants (Nielsen & Rendall, 2011; Fort, Martin, & Peperkamp, 2014) or the presence or absence of corners in a visual shape (Chen et al., 2016). In addition, individuals seem to base sound-symbolic CCs on intrinsic physical properties of shapes and speech sounds.

We also provide nuance to the assumption that the sound-symbolic judgments of pseudowords are based on a conception of rounded and pointedness that derive from orthography of the pseudowords (Cuskley et al., 2015) or in the shape of the lips when producing certain sounds (Ramachandran & Hubbard, 2001; Namy & Nygaard, 2008). Sound-symbolic CCs are not just supported by a conception of shape that links the visual to the auditory domain through production of the sounds or the shape of their written form. Our findings show that sound-symbolic CCs are also related to fundamental physical properties of the visual and auditory stimuli that evoke such CCs.

Visual and acoustic parameters of shapes and pseudowords are of course closely interwoven with the visual and phonetic categories into which the shapes and pseudowords can be divided (Nielsen & Rendall, 2011. Vowel frontness and lip rounding are associated with the frequencies of the first and second formants. Consistent with Nielsen and Rendall's (2011; see also Fort, Martin, & Peperkamp, 2014) finding that lip rounding and vowel frontness are important predictors of judgments of roundedness in pseudowords, Knoeferle et al. (2017) found that the relative frequency of the second formant was an important predictor of judging both the size and the roundedness of shapes. We have evidence that supports Knoeferle et al. (2017) in that the physical parameters of sound are particularly vital to the judgment of roundedness/pointedness of pseudowords, especially from our measurement parameters like the temporal FFT that capture multiple acoustic dimensions of a speech sound. In our results, energy in higher frequencies in the temporal FFT were associated with pointedness, and consistent with the formant analyses of Knoeferle et al. (2017), energy in lower frequencies were associated with roundedness. However, to the extent that physical properties and phonetic features and visual categories are linked, we cannot at present resolve the question of whether sound-symbolic CCs originate from low-level acoustic and visual properties or whether they stem from more general features of phonetic categories.

Our study represents an initial foray into testing the role of multiple visual and acoustic parameters in sound-symbolic CCs. However, our investigation of visual and acoustic parameters was not intended to be exhaustive, and there may be other visual and acoustic parameters not explored here that are important for the perception of sound-symbolic CCs. In particular, we did not find visual or acoustic parameters whose RDMs were significantly negatively correlated with the RDMs of the corresponding perceptual judgments. However, we did find that shimmer, which indicates increased hoarseness or breathiness in the voice (Teixeira & Fernandes, 2014) was negatively correlated with ratings of roundedness. In contrast, the mean HNR was positively correlated with roundedness ratings, and mean HNR is an index of noisiness. Higher HNR indicates less noise and was correlated with higher ratings of roundedness, which is consistent with the fact that lower mean HNR indicates increased hoarseness in the voice (Teixeira & Fernandes, 2014).

In addition, our visual shapes were simplistically constructed to have identical grayscale contrasts and lacked internal visual patterns, which limited the complexity of visual properties that we could sensibly investigate, compared to others with more variable sets of images (Jonas, Spiller, & Hibbard, 2017). The simplicity of our shapes may explain why spatial FFT was not a significant contributor to the perceptual judgments of shape (Petrou & Petrou, 2011). Future studies should examine complex images or examine visual properties important for other types of CCs (Chen et al., 2016; Jonas, Spiller, & Hibbard, 2017).

It is interesting that acoustic properties associated with variation in voice quality, such as shimmer and mean HNR (Teixeira & Fernandes, 2014), were positively correlated with the perceptual judgments. One possible explanation is that the speaker who recorded the stimuli produced pseudowords that were expected to be perceived as rounded or pointed differently, implicitly varying these acoustic properties (Marcel, 1983). However, shimmer and mean HNR reflect properties of speech that are less likely to be affected by bias, as opposed to a more easily altered parameter such as pitch, so the explanation of an unconscious bias is unlikely to completely account for the correlations of these acoustic properties with sound-symbolic judgments. Prosodic cues that speakers are able to modulate can impart meaning of perceptual details in the visual modality, such as color brightness (Tzeng et al., 2018). Sound-symbolic CCs may similarly be affected by speakers' intentional or unintentional changes in vocal patterns to impart meaning (Parise & Pavani, 2011). Future experiments that intentionally vary acoustic properties such as shimmer, mean HNR, and speech envelope within an extensive stimulus set may shed more light on the relative influence of these acoustic properties on sound-symbolic CCs.

Neuroimaging studies of sound-symbolic CCs may further elucidate the nature of the relationship between sound and meaning in these types of stimuli. Currently, there are few published neuroimaging studies in the area of sound symbolism. Perhaps in the presence of sound-symbolic CCs, individuals engage areas of the brain linked to the perception of speech and multisensory integration, like the posterior superior temporal sulcus (Beauchamp, Nath, & Pasalar, 2010). Alternatively, perhaps the presence of sound-symbolic CCs would differentially engage participants' frontoparietal areas that are linked to attention (McCormick et al., 2018; Peiffer-Smadja & Cohen, 2019) and abstract categorization (Revill et al., 2014). Revill et al. (2014) showed that left superior parietal lobule exhibited heightened activation during presentation of sound-symbolic versus non-symbolic words, while fractional anisotropy in the superior longitudinal fasciculus, an area implicated in phonological processing and multisensory facilitation (Brang et al., 2013), correlated with performance of a sound-symbolic task. These authors interpreted their findings as supporting the idea that sound symbolism is processed in areas important for multisensory integration.

2.5.2 Sound symbolism in language acquisition

Neuroimaging methods are one type of approach to the ubiquitous question of why sound-symbolic CCs exist in natural language and how they arise (Sidhu & Pexman, 2017). One explanation to this question is that these associations are a necessary part of language acquisition (Imai & Kita, 2014). Researchers investigating the role of sound symbolism have found evidence to suggest that sound-symbolic CCs are used to determine meanings of words based on individual properties of words and on the broader category to which a word belongs. On the one hand, sound-symbolic CCs provide a tool for children to acquire language by learning the meaning of individual words. Imai et al. (2015) found that the sound structure of individual words bolstered children’s abilities to learn the meaning of that word (see also Ozturk, et al., 2013; Tzeng, Nygaard, & Namy, 2017). This theory is known as the bootstrapping hypothesis (Imai & Kita, 2014). The evidence for this theory indicates that sound-symbolic CCs could be based in individual characteristics of words, at least for young children. Gasser (2004) used computer simulations to illustrate the advantage of sound symbolism in small vocabularies, and this advantage was supported by Brand et al.’s (2018) study of adults learning the individual meanings of smaller sets of pseudowords.

Studies of sound-symbolic CCs also show that they are important for determining meaning across general categories. The findings from Gasser (2004) also indicate that the advantage of determining the specific meanings of words using sound symbolism is diminished as the vocabulary increases (although see Revill et al., 2018). Whether sound symbolism facilitates learning categories or specific word meanings in adults, the idea that sound symbolism can benefit adult learners has been recently validated with eye-tracking and other behavioral studies (Brand et al., 2018; Revill et al., 2018).

Regardless of whether or not sound-symbolic CCs originate from aligned percepts across sensory modalities, having a motivated connection between words and meaning has been suggested to facilitate language learning. The pervasiveness of sound symbolism across languages supports the theory that it may have an important role in language learning (Imai et al., 2015; Kantartzis, Imai, & Kita S, 2011; Tzeng, Nygaard, & Namy, 2017). Individuals are even able to correctly assign meanings of synonym/antonym pairs above chance for languages with which they are unfamiliar (Kunihara, 1971; Nygaard et al., 2009; Tzeng, Nygaard, & Namy, 2016). Our study suggests that one potential motivated connection between sound and meaning is the set of acoustic properties to which listeners are sensitive and that translate systematically into a set of visual properties. If we consider the case of young children learning words, this set of acoustic properties should be connected crossmodally to a set of visual properties in order for the bootstrapping hypothesis to be important for concrete meanings such as shape or size. Sound and meaning may be mapped not only at the word or phonetic level, but perhaps also at the more finely tuned level of acoustic properties and in the case of concrete, physical meanings, also at the level of visual properties. It is also interesting that iconic words such as “moo” and “splash” lead to better performance than non-iconic words on reading aloud and auditory lexical decision tasks in aphasic individuals (Meteyard et al., 2015).

We here provide a key link between continuous acoustic properties and continuous visual properties and show that the perception of roundedness and pointedness along a spectrum in the visual and auditory domains is correlated with similar variations along a spectrum in the simple matching coefficient for visual shapes and the FFT for auditory pseudowords. This finding supports the crossmodal theory that sound symbolism resulted from links across a continuous spectrum for certain auditory and visual properties.

Future work could extend the current study by exploring additional CCs, such as mappings between sound and size (Jonas, Spiller, & Hibbard, 2017), taste (Wang et al., 2016) or emotion (Aryani et al., 2018). It would also be of interest to investigate whether children of different ages or elderly adults are sensitive to similar acoustic and visual properties or if sound-symbolic CCs correspondences change across the lifespan. Studies such as these would further illuminate the potential underlying perceptual and neural mechanisms of sound-symbolic CCs.

2.6 Conclusion

In sum, the current findings suggest that sound-symbolic CCs are driven by distinct acoustic and visual parameters. Specifically, pairwise dissimilarities between the simple matching coefficient, silhouette, and Jaccard distance were correlated with pairwise dissimilarities between judgments of the roundedness/pointedness of visual shapes, while pairwise dissimilarities between the shimmer, FFT, mean HNR, speech envelope, and spectral tilt were correlated with pairwise dissimilarities between sound-symbolic judgments of the roundedness/pointedness of auditory pseudowords. Individuals are not only sensitive to sound-symbolic CCs but are also able to employ visual associations (i.e., rounded vs. pointed) for pseudowords even in the absence of the physical shapes that the sounds may relate to. These findings suggest that sound symbolism is not only a set of visual categorical contrasts that are instantiated by phonetic properties of words, but that individuals are also able to base their sound-symbolic judgments on a continuum of basic visual and auditory properties. Taken together, these findings imply that the relationship between words and their meanings can have a non-arbitrary basis at the level of auditory and visual stimulus properties, and highlight the importance of sound symbolism in natural language.

Chapter 3

Neural Underpinnings of Sound-symbolic Crossmodal Correspondences

3.1 Abstract

One of the fundamental properties of language is the assumed arbitrariness of the relationship between sound and meaning, but sound symbolism refers to non-arbitrary mapping between the sound of a word and its meaning. One way that sound symbolism has been studied is through crossmodal correspondences (CCs) between auditory pseudowords (e.g. ‘loh-moh’) and visual shapes (e.g. a blob). Although sound-symbolic crossmodal correspondences have been studied for almost a century, few have systematically examined the neural regions important for processing sound-symbolic CCs in both the auditory and visual modalities. Here, we used functional magnetic resonance imaging (fMRI) and both univariate and representational similarity analysis to examine the relationships between perceptual ratings that varied on dimensions of roundedness and pointedness, for a range of auditory pseudowords and visual shapes and the neural response to these stimuli. Our analysis indicates that the pattern of perceptual ratings of these stimuli correlate across visual and auditory domains. Contrasting trials in which a multisensory matching (congruent) was detected over trials in which a multisensory mismatch (incongruent) was detected indicates a region of interest in the left precentral-postcentral gyrus may be important for the crossmodal auditory to visual processing sound-symbolic CCs. Further, based on a correlation of perceptual judgments to unisensory session, the bilateral supramarginal gyrus may be important for the visual processing of sound-symbolic CCs, while the left superior occipital gyrus may be important for the auditory processing of sound-symbolic CCs. These findings suggest that the perception of sound-symbolic CCs is complex and multimodal but that both canonically multisensory regions (e.g. the posterior superior temporal sulcus) and unisensory areas (e.g. the superior occipital gyrus) may be important for the perception of sound-symbolic CCs. This research provides insights into the fundamental nature of CCs and that they might evoke specific interpretations of physical meaning in natural language.

3.2 Introduction

The connections between a linguistic form and meaning, such as a word and its referent, are considered to be fundamentally arbitrary (Hockett, 1977; Jackendoff, 2002; Pinker, 1999; de Saussure, 1959). These arbitrary connections are assumed to be a necessary property of natural language and ensure that language is highly flexible in the meanings it can represent (Gasser, 2004). However, there are significant exceptions to the view of arbitrariness, and these exceptions suggest that the relationship between linguistic representations and their meanings can also be non-arbitrary (Köhler, 1929; Sapir, 1929; Sidhu & Pexman, 2017). These opposing viewpoints have been highlighted in works as early as Plato's *Cratylus* (Ademollo, 2011). Behavioral evidence of non-arbitrary correspondences between linguistic structure and meaning is extensive and further indicates that language users are sensitive to these non-arbitrary correspondences (See Chapter 2; Blasi et al., 2016; Nygaard, Cook, & Namy, 2009; Perniss, Thompson, & Vigliocco, 2010).

These non-arbitrary connections are known as sound symbolism and include a vast set of phenomena from onomatopoeia to Japanese mimetics (e.g. 'kirakira' meaning to shine sparklingly or 'guzuguzu' meaning to procrastinate or dawdle) in which the structure of the word resembles or predicts the meaning or characteristics of the referents. Some examples of sound symbolism are different for different languages, such as phonesthemes (Bergen, 2004), which may suggest different mechanisms or origins of sound symbolism, but other examples are reliable and consistent across languages (Blasi et al., 2016; Tzeng, Nygaard, & Namy, 2016).

Individual listeners are even sensitive to cross-linguistic sound symbolism such that they are able to match unfamiliar synonym/antonym pairs to their correct meanings at levels above chance (Kunihira, 1971; Nygaard, Cook, & Namy, 2009). This evidence in combination with consistent sound-symbolic biases across thousands of languages (Blasi et al., 2016) suggests that sound-to-meaning mappings exhibit some consistency across languages that can be recruited to infer the meaning of unfamiliar foreign or novel words (Nygaard, Cook, & Namy, 2009; Brand et al., 2018) or to facilitate early word learning (Imai et al., 2015; Ozturk et al., 2013; Tzeng, Nygaard, & Namy, 2017).

One way to study such a vast set of phenomena as sound symbolism is through sound-symbolic crossmodal correspondences. Crossmodal correspondences are consistent associations between seemingly unrelated features in different sensory modalities, and in the case of sound-symbolic crossmodal correspondences these sensory modalities are assumed to be audition and vision (Spence, 2011). A well-known example of sound-symbolic crossmodal correspondences was first described by Köhler (1929,1947) in which individuals associated a curvy, cloud-like shape with the pseudoword “baluma” and associated an angular star-like shape with the pseudoword “takete”. Crossmodal correspondences often occur between stimulus properties that are occur together, which can lead to more efficient processing of sensory information. For example, large and small size in the visual domain are consistently associated with low- and high- pitched sounds, respectively (Gallace & Spence, 2006; Evans & Treisman, 2010), and high and low elevation are associated with high and low pitch, respectively, in the auditory domain (Ben-Artzi & Marks, 1995; Lacey et al., 2016; Jamal et al., 2017).

One theory of the underlying mechanisms of sound symbolism, particularly cross-linguistic sound symbolism, is based in sound-symbolic crossmodal correspondences. Sidhu and Pexman (2017) reviewed several theories of sound symbolism, one of which is the theory that common patterns of neural coding representing different sensory modalities may lead to sound-symbolic crossmodal correspondences. This theory was proposed by Ramachandran and Hubbard (Hubbard, Brang, & Ramachandran, 2011; Hubbard & Ramachandran, 2005; Ramachandran & Hubbard, 2001), and posits that sound symbolism is a linguistic by-product of multisensory mapping, such that the motor engagement of speech production or the auditory engagement of speech detection further engage activation for corresponding properties in other sensory modalities like vision. In the case of sound-symbolic crossmodal correspondences, these co-activations could theoretically direct associative learning or attention toward particular physical aspects of the objects to which the words refer.

There is further evidence for the neural mechanisms of sound-symbolic crossmodal correspondences in literature on multisensory integration (Noppeney, 2012), on the relationship between multisensory integration and synesthetic associations (Brang et al., 2011; Brang, Williams, & Ramachandran, 2012), and in the potential motor activations present during subvocal articulation (Pulvermüller et al., 2006). For example, synesthetes show enhanced multisensory integration (Brang et al., 2012) and sensitivity to sound-symbolic CCs (Lacey et al., 2016) and differences in anatomy and functional connectivity of parietal areas important for multisensory integration compared to control participants who do not experience synesthesia (Brang et al., 2011; Brauchli et al., 2017; Colizoli et al., 2017). Another theory proposed by Sidhu & Pexman (2017) is that sound symbolism mimics statistical regularities in the natural world, such that high-pitched sounds are present in the context of high locations and low-pitched sounds present from low locations (Jamal et al., 2017; Parise et al., 2014). Through these statistical regularities (Spence, 2011), sound symbolism may be connected with multisensory integration of auditory and visual features (Ković et al., 2010), and neural activity during the processing of sound-symbolic crossmodal correspondences may co-localize with areas that are differentially activated during multisensory integration. For example, some such areas differentially activated during multisensory integration include areas implicated in audiovisual synchrony like the superior temporal cortex Beauchamp, 2005a,b; van Atteveldt et al., 2007; Stevenson et al., 2010; Marchant et al., 2012; Noesselt et al., 2012; Erickson et al., 2014) or areas implicated in audiovisual spatial congruency like the intraparietal sulcus (Sestieri et al., 2006).

Very little research has been conducted on the neural basis of sound symbolism (Ković et al., 2010; Peiffer-Smadja & Cohen, 2019; Revill et al., 2014; McCormick et al., BioRxiv), and most previous studies have not examined both auditory and visual processing of sound-symbolic crossmodal correspondences (but see McCormick et al., BioRxiv). While some studies have investigated sound-symbolism across a variety of semantic dimensions (Ković et al., 2010; Revill et al., 2014), it has not been established that all sound-symbolic associations are processed in the same ways and from the same neural substrates (Sidhu & Pexman, 2017). Further, electroencephalograph (EEG) studies (Ković et al., 2010; Sučević et al., 2015; Lockwood et al., 2016), provide excellent temporal information, but may not have sufficient anatomical resolution to pinpoint exact brain regions.

In previous work, McCormick and colleagues developed a large systematically varied set of pseudowords and shapes to investigate the acoustic (McCormick et al., 2015; McCormick et al., BioRxiv) and visual parameters (McCormick et al., unpublished data) that drive judgments of sound-symbolic crossmodal correspondences (List et al., BioRxiv; See Chapter 2). In Likert type rating tasks, independent sets of participants rated these pseudowords and shapes (McCormick et al., 2015; McCormick et al., BioRxiv; McCormick et al., unpublished data). Although participants were rating all pseudowords on the same spectrum of roundedness/pointedness, how consistently participants rated a pseudoword as more rounded or pointed varied across individual pseudowords. This variation indicates that our stimulus set captured a wide spectrum of pseudowords from those that were highly sound-symbolic to those that were considered neither rounded nor pointed. From these data, we identified a set of shapes and pseudowords that were judged across the spectrum from pointed to rounded (McCormick et al., 2015; McCormick et al., BioRxiv; McCormick et al., unpublished data).

We used functional magnetic resonance imaging (fMRI) to investigate cerebral cortical localization of congruency effects related to the sound-symbolic crossmodal correspondence between auditory pseudowords and visual shapes. In order to further test the functional significance of the areas identified during the processing of sound-symbolic crossmodal correspondences, we compare across auditory and visual modalities using representational similarity analysis (RSA) (Kriegeskorte et al., 2008). This method relies on representational dissimilarity matrices (RDMs) and has traditionally been applied to capture the structure of dissimilarity for individual stimuli or distinguish classes of stimuli (Kriegeskorte et al., 2008; See Chapter 2). We here use RSA to compare patterns of similarity between judgments of systematically varied shape and pseudoword stimuli and neural data during the perception of these same shape and pseudoword stimuli. RSA allows us to develop a comprehensive view of the commonalities and the differences in the patterns that drive both visual and auditory neural correlates of sound-symbolic crossmodal correspondences.

To address the question of which cerebral cortical areas form the basis of sound symbolism, the present study was designed to systematically determine how auditory, visual, and multisensory perception contribute to the perception of sound-symbolic crossmodal correspondences using RSA. Specifically, we examined whether differential activation in a key set of cortical brain areas reliably predicted sensitivity to sound-symbolic crossmodal correspondences.

In our study, we compare the relative influence of multiple cerebral cortical areas during auditory and visual perception of the component stimuli that elicit sound-symbolic crossmodal correspondences. First, we confirmed participants' sensitivity to sound-symbolic crossmodal correspondences with two independent behavioral tasks, one inside and one outside the scanner. Then, we identified candidate cortical brain regions during the multisensory perception of a sound-symbolic crossmodal correspondence using a univariate approach in which we contrasted congruent (or matching) over incongruent (or mismatching) pairs of stimuli. We also investigated the extent to which the determination of rounded versus pointed in a single sensory modality may be important for the multisensory decision of congruent or incongruent by contrasting rounded and pointed stimuli in visual and auditory unisensory sessions. To better understand the effects of unisensory stimuli to create the multisensory percept of sound symbolic crossmodal correspondences, we also investigated the extent to which perceptual ratings are correlated with activity in unisensory areas. Finally, we performed RSA in these regions on BOLD signal data collected during the perception of visual shapes and separately during the perception of auditory pseudowords to assess the roles of visual perception and auditory perception, respectively, in sound-symbolic crossmodal correspondences.

3.3 Materials and Methods

3.3.1 Participants

Thirty participants took part in this study. Of these participants, six chose not to complete the study referencing an excessive length of the experiment, leaving a remaining total of twenty-four participants who completed all three sessions (14 female, 10 male, mean age = 22.6 years, SD = 3.8). Participants were recruited via flyers placed around Emory University and Georgia Institute of Technology, Savannah College of Design, Agnes Scott College, Gwinnett Technical College, and Georgia State University. All were right-handed based on the validated subset of the Edinburgh handedness inventory (Raczkowski et al., 1974) and reported normal hearing and normal, or corrected-to-normal, vision. All participants gave informed consent and were compensated for their time. All procedures were approved by the Emory University Institutional Review Board.

3.3.2 General

Participants performed three scanning sessions described in detail below, consisting of a unisensory auditory, a unisensory visual, and finally a multisensory session. Participants performed the unisensory auditory scans first, then the unisensory visual scans, and finally the multisensory auditory-visual scans to investigate the cerebral cortical areas underlying the sound-symbolic crossmodal correspondence. After these scans, they performed a behavioral task to determine the strength of their sound-symbolic crossmodal correspondence. Since McCormick and colleagues previously found the selected shape and pseudoword stimuli to be related in judgments of roundedness/pointedness (McCormick et al., 2015; McCormick et al., BioRxiv; McCormick et al., unpublished data; See Chapter 2), this fixed order was followed to avoid potential priming effects of the unisensory visual scans on the unisensory auditory scans and of the behavioral tasks on the multisensory scans. For multivariate analyses, the multisensory scan served as a localizer for the unisensory scans. Each scan comprised twelve runs with a fixed pseudorandom stimulus order within each run; the order of runs was counterbalanced across participants. All participants completed the scans over three sessions, and the inter-scan interval was an average of 80 hours (range 27-136 hours). All experiments were presented via Matlab software (The MathWorks, Inc., Natick, Massachusetts) which allowed synchronization of scan acquisition with experiments and also recorded responses and response times (RTs). Behavioral data was also analyzed in Matlab.

3.3.3 Unisensory visual and unisensory auditory fMRI tasks

From the set of 90 shapes and 537 pseudowords McCormick and colleagues created (See Chapter 2; McCormick et al., 2015; McCormick et al., BioRxiv; McCormick et al., unpublished data), we selected twelve shapes and twelve pseudowords that varied on the scale of pointedness to roundedness based on the average judgements of previous participants (See Figure 3.1). The pseudowords were digitally recorded in a female voice using Audacity v2.0.1 (Audacity Team, 2012), with a SHURE 5115D microphone and an EMU 0202 USB external sound card, at a 44.1kHz sampling rate. The recordings were then processed in Audacity, using standard tools and default settings, edited into separate files, amplitude-normalized, and down-sampled to a 22.05kHz sampling rate. Stimulus duration was an average of 457.5 ms (SD of 62 ms). The visual stimuli were gray outline shapes on a black background (Figure 3.1 bottom), each subtending approximately 1° of visual angle and presented at the center of the screen for 500ms. A mirror angled over the head coil enabled participants to see the visual stimuli projected onto a screen placed in the rear magnet aperture. Auditory stimuli were presented via scanner-compatible headphones.

Figure 3.1 Stimuli presented for unisensory auditory (top) and unisensory visual (bottom) scans are ordered from pointed to rounded, according to separate participants' judgments (See Chapter 2).

There were twelve runs in the unisensory auditory scan and another twelve runs in the unisensory visual scan in two slow event-related designs. Each run contained 26 trials (2 trials for each of the 12 stimuli plus two 'repeated' trials) presented in pseudorandom order with a 10s interval between trials and 6 baseline periods of 10s duration which occurred at the start of a run, every ten trials thereafter, and at the end of a run. Total run duration was 320s. Participants performed a one-back rare repeats task by pressing a button on a hand-held response box. 'Repeated' trials, i.e. the same auditory or visual combination occurring in two consecutive trials, constituted 7.7% of trials in each run.

3.3.4 Multisensory fMRI task

The selected stimuli, two visual shapes and two auditory pseudowords (Figure 3.2) for the multisensory task lay near the ends of independent rounded and pointed dimensions in each modality based on empirical ratings (See Chapter 2; McCormick et al., 2015), i.e., the chosen rounded pseudowords and shapes were rated towards the high end of the rounded dimension and the low end of the pointed dimension, and vice versa for the chosen pointed stimuli. These stimuli were presented concurrently in audiovisual pairs that were either congruent ("kehteh"/pointed shape or "mohloh"/rounded shape; Figure 3.2) or incongruent ("kehteh"/rounded shape or "mohloh"/pointed shape) with respect to the crossmodal pseudoword-shape (sound-symbolic) correspondence.

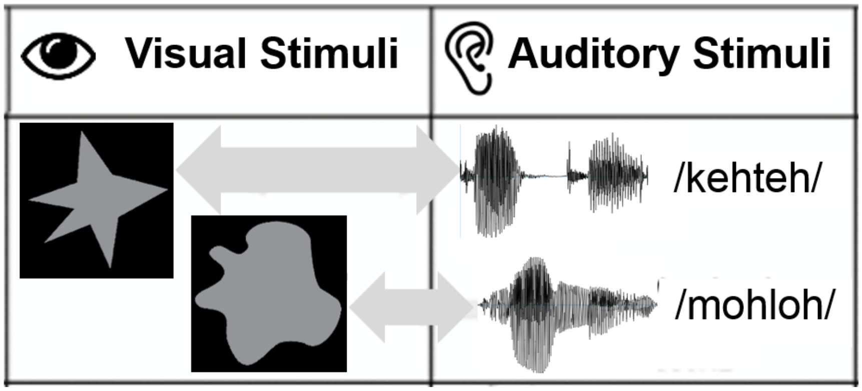


Figure 3.2. Exact stimuli for the multisensory task (shown here in two congruent trial combinations) were selected to elicit sound-symbolic crossmodal correspondences and based on perceptual ratings from previous participants.

In this task, there were twelve runs in a jittered event-related design. To optimize the interstimulus interval, we used optseq2 (Optseq2, 2006), which resulted in the distribution of ISIs in Figure 3.3. Each run consisted of 22 congruent trials and 22 incongruent trials, each lasting 2 seconds (500 ms of stimulus presentation and 1.5 s of time for response) with 6 baseline periods of 10-30s duration which occurred at the start of a run, interleaved according to the optseq2 optimization procedure, and at the end of a run. Total run duration was 326s. Pseudoword-shape pairs were pseudorandomly interleaved (no more than three trials in a row of the same pairing, no more than two trials in a row of the same condition). Participants attended to each auditory-visual stimulus pair and judged whether the pair was a match or a mismatch, pressing one of two buttons on a hand-held response box when they heard and saw either a pair that matched or that mismatched. The right index and middle fingers were used to indicate responses, one finger for each condition, counterbalanced among subjects. All participants except for one judged the ‘match’ condition to be congruent pairs of stimuli and the ‘mismatch’ condition to be incongruent pairs of stimuli (Figure 3.2). For this participant, we remained consistent in comparing response times for congruent versus incongruent stimuli. Thus the hypothesis for this individual would be that incongruent trials (their ‘matching’ condition) would be faster than congruent trials, which were their ‘mismatching condition’.

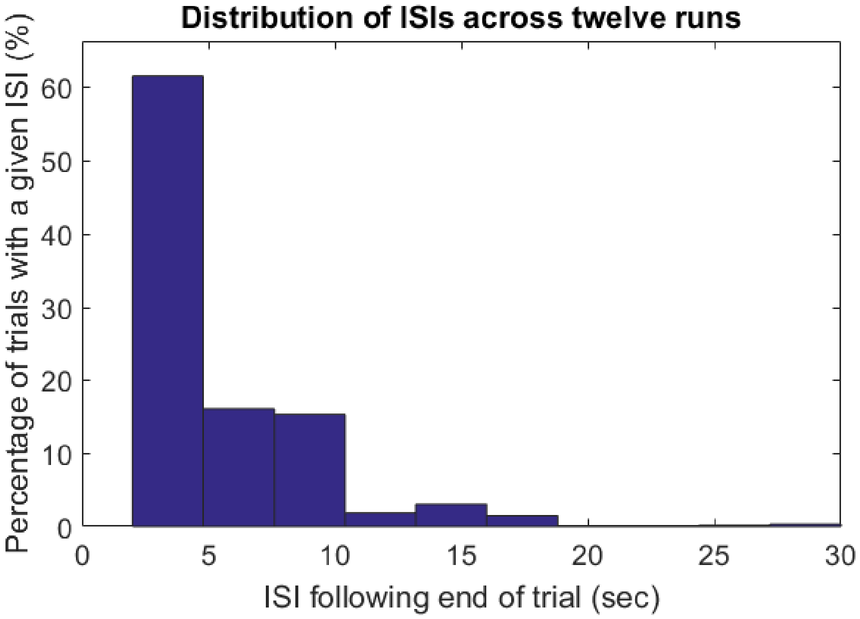


Figure 3.3. Interstimulus interval (ISI) was jittered and optimized according to a minimum interval of 4 seconds and maximum interval of 32 seconds.

3.3.5 Post-scan behavioral testing

As a final step, we tested whether participants reliably demonstrated the crossmodal sound-symbolic judgments using a Likert-type rating task (See Chapter 2). Participants completed two runs of four Likert-type ratings tasks in which they were asked to rate pseudowords or shapes using one of two Likert-type scales of roundedness to pointedness from 1-7 as described in Chapter 2. Pseudowords were presented auditorily over Beyerdynamic DT100 headphones at approximately 75db SPL. For the pseudowords on each trial, the Likert-type scale from 1-7 appeared on the screen. For the shapes on each trial, the Likert-type scale from 1-7 appeared below the shape.

All participants rated pseudowords over four runs and shapes over another four runs for a total of eight runs. Each run consisted of 36 trials for a total of twelve ratings per each of the twelve pseudowords and another twelve ratings per each of the twelve shapes used in the experimental task. Each run was 123.8 seconds long. Trials consisted of a 5 second blank period at the beginning, a 500 ms stimulus presentation, and a 2.8 second ISI during which participants rated the stimulus on a Likert scale of 1-7 using the keyboard. For two runs, participants rated each pseudoword for degree of pointiness with 1 representing not pointed and 7 representing very pointed, and for another two runs, participants rated each pseudoword for degree of roundedness with 1 representing not rounded and 7 representing very rounded. Participants also rated each shape during two runs for degree of pointiness with 1 representing not pointed and 7 representing very pointed, and during another set of two runs, participants rated each shape for degree of roundedness with 1 representing not rounded and 7 representing very rounded. The order of the four Likert-type ratings tasks were counterbalanced across participants, and the response keyboard always had 1-7.

We used these rating results obtained from the post-scan behavioral task to create a reference representational dissimilarity matrix (RDM) based on perceptual ratings (See 3.3.8 Multivariate analysis). Each of the two perceptual rating RDM models, one visual shape and one auditory pseudoword, were constructed at the individual subject level by calculating pairwise Pearson correlations between the ratings for different shape or pseudoword stimuli, respectively. The perceptual rating reference RDMs were computed separately for each subject and for visualization purposes were then averaged across subjects.

3.3.6 Image acquisition

MR scans were acquired in the same manner using the procedure outlined by McCormick et al. (2018) The main two differences included number of scans acquired per run and the programs used to run the experimental tasks. For the multisensory session, there were 163 acquisitions per run, and for the unisensory sessions there were 160 acquisitions per run. Also, once magnetic stabilization was achieved in each run, the scanner triggered the computer running Matlab 2014b (The MathWorks, Inc., Natick, Massachusetts) software and the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007) so that the sequence of experimental trials was synchronized with scan acquisition. For additional details, see section 2.3 Image acquisition in McCormick et al. (2018).

3.3.7 Image processing and univariate analysis

Image processing and analysis was performed using SPM12 (Wellcome Trust Center for Neuroimaging, London, UK). In individual analyses, each participant’s functional runs were real-time motion-corrected utilizing Siemens 3D-PACE (prospective acquisition motion correction) during acquisition, so post-scan motion parameters were not included as a covariate in analysis. Functional images were preprocessed utilizing slice scan time correction, 2˚ B-spline interpolation for intra-session alignment of functional volumes, and high-pass temporal filtering of 128 seconds to remove slow drifts in the data without affecting task-related effects. Anatomic 3D images were processed, co-registered with the functional data, and transformed into MNI space (Evans et al., 1993; Mazziotta et al., 1995). MNI-normalized anatomic data sets from multiple scan sessions (3 per participant) were averaged for each individual, to minimize noise and maximize spatial resolution.

For univariate group analyses in each session, the MNI-transformed data were 3D spatially smoothed with an isotropic Gaussian kernel (full-width half-maximum 8mm). The 8mm filter is within the recommended limit of slightly larger than the original voxel size of 3 mm to reduce the possibility of blurring together activations that are in fact anatomically and/or functionally distinct (Pajula & Tohka, 2014). In accordance with the procedure of McCormick et al., (2018), runs were percent signal change normalized (i.e., the mean signal value for each voxel's time course was transformed to a value of 100, so that the individual values fluctuated around that mean as percent signal deviations).

For group activation display and statistical analyses, we followed the procedure described by McCormick et al. (2018) with the exceptions of using 4˚ B-spline interpolation following co-registration and using SPM12, which uses MNI instead of Talaraich space. This procedure resulted in a 24-subject MNI template was then manually segmented, creating a group average cortical ‘mask’ file with 3mm spatial resolution, equivalent to the spatial resolution of the functional data files.

Statistical analyses of group data used general linear models (GLMs) treating participant as a random factor (so that the degrees of freedom equal n-1, i.e. 23), followed by pairwise contrasts. This analysis allows generalization to untested individuals. Correction for multiple comparisons within a cortical mask ($p<0.05$, FWE corrected or $p<0.001$ uncorrected where noted) was achieved in SPM12 using the FWE procedure based on Gaussian random field theory (RFT, Han & Glenn, 2017). Activations were localized with respect to 3D cortical anatomy with the help of an MRI atlas (Duvernoy, 1999) and the SPM extension AAL2 (Automated Anatomical Labeling, Tzourio-Mazoyer et al., 2002; Rolls et al., 2015).

For the purposes of the univariate analyses, the multisensory session was used as an initial result by which to understand how sound-symbolic CCs are perceived in the multisensory condition. We used the contrast of congruent>incongruent (See 3.3.4 Multisensory fMRI task) to determine which cerebral cortical areas were important for the perception of the sound-symbolic association of “pointed” pseudowords with pointed shapes and “rounded” pseudowords with rounded shapes. Then, we were interested in to what extend the distinction of pointed or rounded in a single modality may influence the decision to make a sound-symbolic association across sensory modalities. To determine the differences or overlap of cerebral cortical areas in the perception of unisensory sound-symbolic stimuli compared to the full multisensory sound-symbolic CC, we contrasted “pointed” and “rounded” pseudowords and pointed and rounded shapes, respectively. Finally, we were interested in understanding the manner in which post-scan perceptual ratings may be correlated with activation in the unisensory conditions. We conducted an item analyses in which each stimulus>baseline activation was correlated with the average perceptual rating for that stimulus to understand how the ratings of pointedness/roundedness of a stimulus may be correlated with cerebral cortical areas important for perceiving such a unisensory stimulus. We are here questioning whether the unisensory distinctions among these stimuli that contribute to their perceptual ratings are related to the percept of sound-symbolic CCs.

3.3.8 Multivariate analysis

For purposes of the multivariate analyses, we used the multisensory session as a localizer in which we identified multiple candidate regions of interest (ROIs). We then performed a multivariate analysis called representational similarity analysis (RSA) on these ROIs using the data from the unisensory sessions to determine how the potential multisensory regions that we previously identified may be important for sound-symbolic CCs in a unisensory context. For each ROI and each participant, we performed RSA for the t-statistic of the BOLD response comparing each stimulus to baseline. Once we obtained one RDM for each ROI for each participant for the data obtained while they were viewing shapes, we compared that RDM to the RDM from their rounded/pointed ratings of those same visual shapes using a second order Spearman correlation. We performed the same procedure for the auditory stimuli for a total of 912 neural RDMs across modality (2), participant (24), and ROI (19).

To identify the areas that were sensitive to sound-symbolic crossmodal correspondences at the group level, we performed a second level GLM for the multisensory session and defined regressors of interest corresponding to congruent and incongruent trials using the canonical HRF.

For the areas derived from this contrast (See 3.4 Results), we drew a 7 mm-radius sphere using the MarsBar toolbox (version 0.44, Brett et al. 2002) around the coordinate that was maximally activated in a congruent > incongruent contrast in SPM. For each visual shape and auditory pseudoword stimulus condition, we extracted the average voxel time series in the multisensory ROIs and calculated the percent BOLD signal change using MarsBar for multivariate analyses.

The goal of representational similarity analysis (RSA) is to test explicit hypotheses about how a feature of interest (i.e., seeing shapes of varying rounded/pointedness) is represented in distributed activity patterns in an ROI. RSA provides a measure of the spatial distribution of responses that is stable across different levels of overall activation (Kriegeskorte et al., 2008; Arbuckle et al., 2018). For both the unisensory visual and unisensory auditory sessions, contrasts of stimulus greater than baseline were defined from a first-level GLM analysis for each shape or pseudoword event for each functional run. We used the resulting *t*-values from the contrasts as input for the representational dissimilarity matrices (RDMs), constructed using the Decoding Toolbox (Hebart et al., 2015) and custom Matlab scripts. We used the Mahalanobis distance to estimate the dissimilarity between multivariate voxel response patterns within an ROI for each pair of target size conditions. Each pairwise distance estimate (e.g., ‘mohnoo’ v. ‘keekeh’) corresponds to one cell in the RDM. The Mahalanobis distance is a more reliable way to calculate RDMs than correlation metrics or Euclidean distance because it accounts for differing noise levels across voxels and spatially correlated noise in nearby voxels (Walther et al., 2016). This weighting scheme is implemented by estimating the noise-covariance matrix from the residuals of the GLM for each run.

BOLD RDMs were computed separately for each run and were then averaged across runs and subjects. To further infer the representational structure of the multisensory ROIs, we performed a second-order Spearman correlation between the neural RDMs and the perceptual rating model (See 3.3.5 Post-scan behavioral testing). The second-order correlations were Fisher z-transformed and then averaged across subjects. Mean correlations were corrected for multiple comparisons using the false discovery rate correction (Benjamini & Hochberg, 1995; Genovese et al., 2002).

3.4 Results

3.4.1 Behavioral

3.4.1.1 In-scanner tasks

In the unisensory auditory task, participants correctly responded (mean ± sem) to 89.2±1.2% of the rare repeat trials on average. In the unisensory visual task, participants correctly responded to 85.3±1.9% of the rare repeat trials on average. For the multisensory task, participants correctly identified the congruency of the match/mismatch task 93.3±0.7% of the trials on average. Mean (±sem) accuracy was not significantly different between trials requiring an incongruent response (93.1±1.21%) and those requiring a congruent response (93.5±0.67%: $t_{23} = 0.51$, $p = .6$). The false alarm rates for both congruent and incongruent trials were quite low (2.5% vs. 2.7% respectively) and hence we did not attempt to correct for response bias by computing d′. RT analyses excluded trials for which there was no response (1.3% of all trials), incorrect responses (5.3% of responses), and further excluded trials for which the RT was ± 2.5 standard deviations from the individual participant mean (3.1% of correct response trials). RTs were significantly different between congruent (1101±42ms) and incongruent (1199±53ms) trials ($t_{23} = 5.03$, $p<0.001$) with an average difference of 79 ms±16 ms. Four participants failed to show a congruency effect, i.e. faster RTs for congruent trials, and a fifth participant mapped congruent trials as “mismatch” and vice versa for incongruent trials. For this participant, we still compared times for congruent versus incongruent stimuli. Thus for this participant, the incongruent trials, which were their ‘matching’ condition, were faster than congruent trials, which were their ‘mismatching’ condition. Because this participant made an assessment opposite to that of most participants, perhaps they are integrating the stimulus pairs that most do not, leading to a faster response for canonically incongruent trials compared to canonically congruent trials. All participants were included in our analyses. We computed a normalized measure of congruency with the following equation: (Incongruent RTs – Congruent RTs)/(Incongruent RTs + Congruent RTs). The average normalized congruency was 0.043±0.008, which is significantly greater than zero ($t_{23} = 5.18$, $p < 0.001$) and shows that this group of participants was sensitive to sound-symbolic crossmodal correspondences in the scanner.

3.4.1.2 Post-scan behavioral testing

Here we used RDMs for perceptual ratings, and the group RDM was averaged across participants as noted in the methods (3.3.8). The pattern evident in the 12x12 RDM for the pointed/roundedness ratings for the visual shapes (Figure 3.4, left) was largely binary. This pattern indicates that participants rated shapes largely as more pointed or more rounded, with only a few shapes considered intermediate. Figure 3.4 (right) depicts the 12x12 RDM of pointed/roundedness ratings for the pseudowords, which appeared to be more graded. The two matrices for the shapes and pseudowords were positively correlated (Spearman $r = 0.63$, $p<0.0001$).

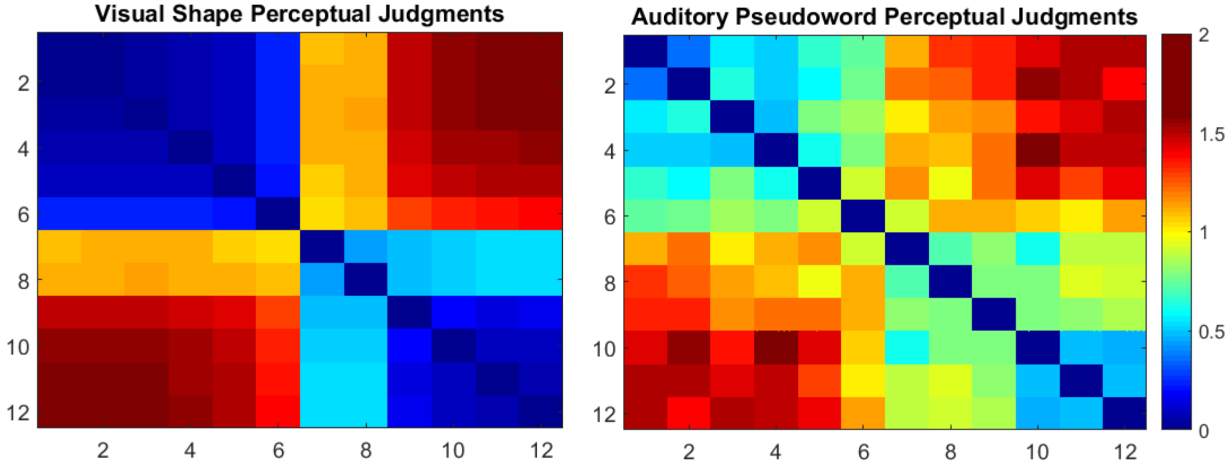


Figure 3.4. RDMs ordered according to mean perceptual judgments for visual shapes (left) and auditory pseudowords (right). Based on the pattern of correlations, visual shape judgments tended to be binary. These results replicate our previous findings in larger stimulus sets (See Chapter 2). Although some pseudowords were judged to be intermediate, there was a significant positive correlation between the two RDMs (Spearman r = 0.63, $p<0.0001$).

This finding indicates that the judgments were crossmodally consistent. This finding is to be expected given that participants were asked to rate stimuli along dimensions of pointedness or roundedness in both modalities and that the stimuli were selected from a larger set for which the same task was performed in independent sets of participants, but this finding replicates our original findings in Chapter 2 with a subset of stimuli and using a within participant design. The strong positive correlation between the auditory and visual RDMs establishes the presence of sound-symbolic CCs between a subset of 12 auditory pseudowords and 12 visual shapes from the original 537 auditory pseudowords and 90 shapes. This perceptual reference lays the foundation for our subsequent analyses of the cortical brain regions that are important for the processing of sound-symbolic CCs.

3.4.2 Imaging

3.4.2.1 Univariate Analyses with Multisensory fMRI Task Data

For a list of directional and anatomical/functional abbreviations, see Table 3.1 or the Abbreviations section. The contrast of congruent > incongruent within the cortical mask (FWE-corrected $p < .05$, cluster threshold 5 voxels) revealed multiple cerebral cortical areas that may be important for the multisensory processing of sound-symbolic CCs, at least in the condition when the stimuli are integrated and congruent (Table 3.2). These cerebral cortical areas were then used for RSA (Figure 3.10-3.12). Based on the findings of Noppeney (2012) and Walker (2016) that posit that the process of failing to integrate during the incongruent condition can lead to increased cortical activation, we were motivated to examine the reverse contrast of incongruent > congruent, to explore the possibility that this contrast would yield differential activations as a result of mismatching sensory inputs. This contrast, within the cortical mask (FWE-corrected $p < .05$, cluster threshold 5 voxels), revealed no significant activations. We further examined our behavioral measures as covariates, including normalized congruency and second order Spearman correlation for the perceptual judgments, but these maps, even with an uncorrected threshold (voxel-wise threshold $p < .001$, cluster threshold 5 voxels) were also blank.

Abbreviation	
(a)	
a	anterior
d	dorsal
L	left hemisphere
med	medial
p	posterior
R	right hemisphere
v	ventral
(b)	
AG	angular gyrus
CP	caudate putamen
FFG	fusiform gyrus
Ins	insula
IOG	inferior occipital gyrus
MFG	middle frontal gyrus
MOG	middle occipital gyrus
MTG	middle temporal gyrus
preCG	precentral gyrus
postCG	postcentral gyrus
PO	parietal operculum
SMA	supplementary motor area
SMG	supramarginal gyrus
SOG	superior occipital gyrus
VWFA	visual word form area
V1	primary visual cortex

Table 3.1. Directional (a) and anatomical/functional (b) abbreviations for results of neuroimaging. All other abbreviations are in the text or see Abbreviations section.

The univariate contrast of congruent>incongruent revealed multiple regions that are known to be important for auditory processing, visual processing, auditory-visual processing, and speech processing (Table 3.2). Listed first are the areas from the left hemisphere in the order of frontal, parietal, temporal, occipital lobe areas and then basal ganglia and cerebellar areas, and second listed are the areas from the right hemisphere in same anatomical order. The large cluster in the left hemisphere included the preCG and postCG (Figure 3.5 for visualization). The multisensory task involved right-handed button press, but the contrast of congruent>incongruent should have subtracted the motor response unless that motor response was somewhat different (e.g. faster) for congruent versus incongruent responses. Alternatively, motor cortex has been shown to map articulatory features of speech sounds (Pulvermüller et al., 2006). This cluster also extended into areas that have been implicated in subvocal articulation (Price et al., 2009). In the left hemisphere, our analysis revealed the posterior superior temporal gyrus (p STG), two clusters in the middle temporal gyrus (a MTG), the fusiform gyrus (FFG) that may possibly be labelled the visual word form area (VWFA) based on the coordinates in Cohen et al. (2002), and the parietal operculum (PO) that are sensory areas. This contrast also captured three clusters in the supplementary motor area (SMA), which may have been due to the task that was completed using the contralateral (right) hand, but that effect should have been subtracted out from the contrast. Left FFG has been shown to be important in reading and language processing (Pulvermüller et al., 2006). AG has been implicated in auditory-visual integration in TMS studies (Chambers et al., 2007; Kamke et al., 2012). Areas such as the pSTG, the MTG, and the PO, have been discussed as important for multisensory integration, and in the case of the PO, may be playing a role in attention during MSI (Basso et al., 2012; Bolognini et al., 2010; Perrodin et al., 2014; See Chapter 1, Section 1.2.2). While pSTS has been shown to mediate multisensory processing in a speech illusion (Beauchamp, Nath, & Pasalar, 2010; McGurk et al., 1976), the pSTG has been shown to mediate categorical speech perception such that the neural representation of speech sounds is divided among phonetic categories across the pSTG (Chang et al., 2010)

More clusters appeared on the left hemisphere than on the right hemisphere (Table 3.2), which may be explained by the fact that sound-symbolic CCs involved speech sounds, and many neural correlates of language are left lateralized in the brain (Pulvermüller et al., 2006). However, from a univariate contrast of congruent > incongruent, several areas important for sensory processing appeared including the middle, superior, and inferior occipital gyri (MOG, SOG, and IOG), the caudate and putamen (CP), and two clusters in the supramarginal gyrus (SMG; Chambers et al., 2007; Romei et al., 2009; Vetter et al., 2014; See Chapter 1, Section 1.2.2). The multiple occipital regions from our results may be functioning purely as visual processors or they may act with areas such as the SMG to modulate auditory-visual multisensory integration (Romei et al., 2009; Lewis & Noppeney, 2010).

Region	**Cluster Size**	**x**	**y**	**z**
L Middle Frontal Gyrus	7	-32	40	22
L Precentral Gyrus-Postcentral Gyrus	1601	-48	-16	30
	35	-4	-20	54
L Supplementary Motor Area	109	-8	0	46
L Medial Cingulate Sulcus	15	0	-32	56
L Parietal Operculum	5	-42	-30	22
L Ventral Angular Gyrus	13	-36	-46	34
L Posterior Superior Temporal Gyrus	989	-48	-42	20
L Anterior Middle Temporal Gyrus	21	-56	-32	-8
L Posterior Middle Temporal Gyrus	5	-52	-28	-12
L Fusiform Gyrus	7	-38	-50	-18
R Postcentral Gyrus	84	52	-20	54
R Anterior Insula	6	32	18	8
R Anterior Supramarginal Gyrus	36	66	-42	26
R Posterior Supramarginal Gyrus	6	36	-38	46
R Superior Occipital Gyrus	5	22	-82	38
R Middle Occipital Gyrus	283	46	-78	24
R Inferior Occipital Gyrus	9	48	-78	-4
R Caudate Putamen	32	26	-4	4

Table 3.2. Multisensory fMRI session activations: congruent > incongruent, within cortical mask, FWE-corrected p < .05, cluster threshold 5 voxels; x, y, z: MNI coordinates for local maximum (Woo et al., 2014). These areas are candidates for RSA in the unisensory visual and unisensory auditory tasks.

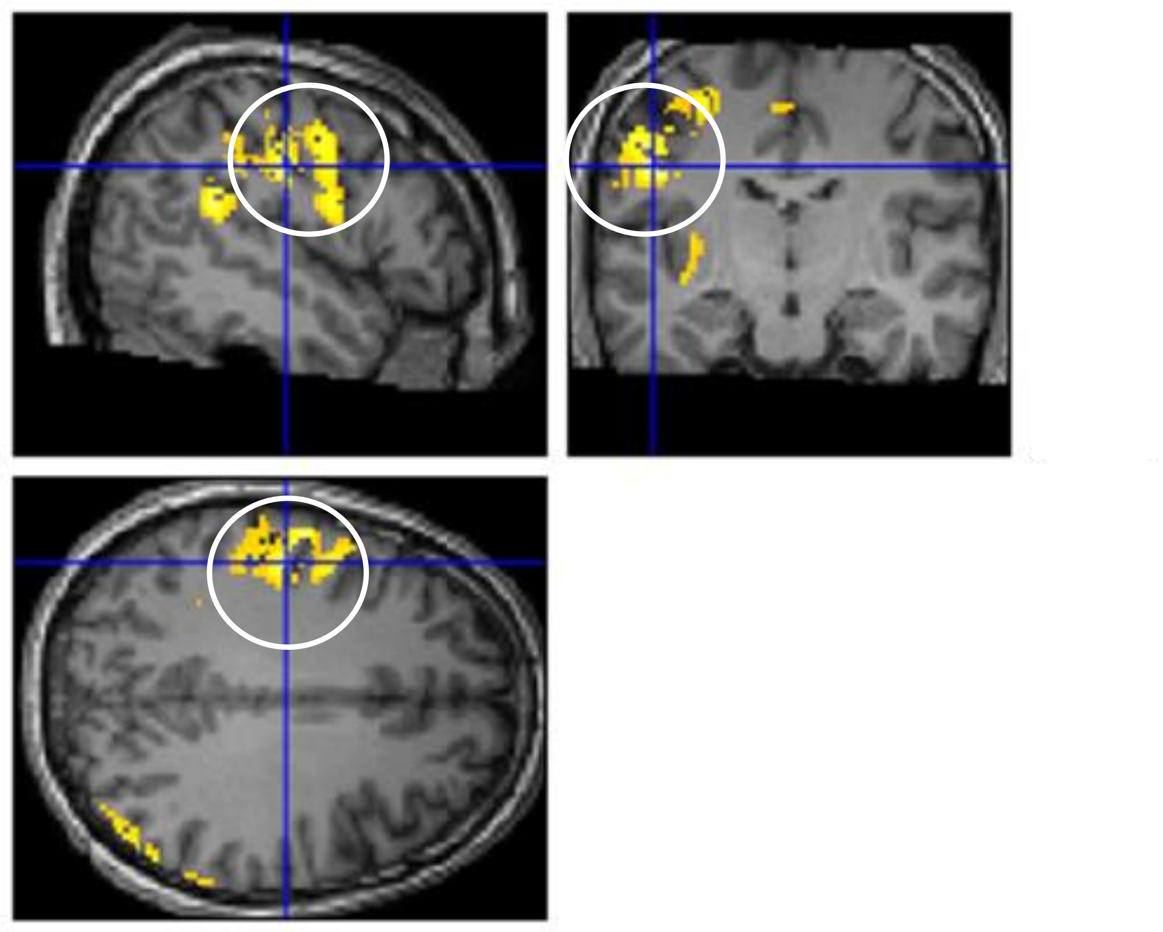


Figure 3.5. View of the left precentral gyrus-postcentral gyrus cluster (cluster size 1601 voxels circled in white) at the second level group analysis (n = 24; yellow) on the multisensory session with a univariate contrast of congruent > incongruent (FWE $p < 0.05$). Panels include a sagittal (top left), coronal (top right), and transverse (bottom left) view. Within the transverse panel, right is anterior, left is posterior. Within the coronal panel, left is left hemisphere, and right is right hemisphere. The large and diffuse area which appears via the congruent>incongruent contrast is a cortical area that may be important for the multisensory perception of sound-symbolic CCs and is a potential candidate for RSA.

In sum, the univariate contrast of congruent > incongruent while participants perceived sound-symbolic CCs in the scanner revealed multiple areas, primarily in the left hemisphere but also in the right, that may be important for the processing of sound-symbolic CCs, for speech or language processing, for sensory processing, and even auditory-visual multisensory integration more generally.

3.4.2.1 Univariate Analyses with Unisensory Visual and Unisensory Auditory fMRI Task Data

After investigating the multisensory results, we next examined the unisensory visual and unisensory auditory tasks with univariate analyses. To determine the extent to which congruency is determined via the distinction of rounded versus pointed in each unisensory modality, we contrasted pointed>rounded in the auditory session We divided the twelve stimuli in each respective modality, shapes for visual and pseudowords for auditory, such that six stimuli were categorized as a “pointed” event and the other six were categorized as a “rounded” event. The contrast of pointed > rounded in the visual session within the cortical mask (FWE-corrected $p < .05$, cluster threshold 5 voxels) revealed multiple additional areas that may be important for the perception of sound symbolic crossmodal correspondences in the visual modality (Table 3.3). Unlike in the multisensory analyses, areas such as the left inferior frontal gyrus (IFG), the frontal operculum (FO), several cerebellar clusters, and the anterior intraparietal sulcus (IPS) are present, as well as areas that are important for language processing, such as the left supramarginal gyrus (SMG). As in the multisensory analyses, we see insula and supramarginal gyrus in the pointed > rounded contrast for the visual session, but in the left hemisphere instead of the right. We also see pre and post central gyrus in the left hemisphere for both multisensory congruent>incongruent and unisensory visual pointed > rounded. Surprisingly, we do not see areas that were present in the multisensory analyses that are visual areas, such as the right IOG, SOG, and MOG.

In the visual session, we also examined the reverse contrast of rounded > pointed within the cortical mask and found that this contrast was somewhat less robust (uncorrected $p < .001$, cluster threshold 4 voxels), but revealed several important areas (Table 3.4). Like the pointed>rounded contrast for the visual session, we see LG, but in the right hemisphere instead of the left. In this contrast, we also observe middle cingulate and importantly, the visual cortical area V1 for the distinction of rounded over pointed stimuli.

Region	**Cluster Size**	**x**	**y**	**z**
L Inferior Frontal Gyrus	1	-36	8	30
L Opercular Part of the Inferior Frontal Gyrus	1	-50	6	32
	1	-56	16	24
L Anterior Intraparietal Sulcus	4	-36	-34	40
L Precentral Gyrus	1	-40	-12	30
	1	-28	-28	54
L Ventral Precentral Sulcus (Premotor Cortex)	1	-28	-10	54
L Postcentral Gyrus	1	-38	-30	36
	3	-40	-18	46
	1	-46	-16	52
	1	-46	-14	28
L Postcentral Sulcus	1	-46	-32	46
L Supramarginal Gyrus	2	-50	-30	42
	1	-42	-30	36
L Angular Gyrus	1	-48	-66	38
L Posterior Insula	1	-34	-32	14
	1	-36	-12	-8
L Lingual Gyrus	1	-8	-54	-8
L Anterior Cerebellum	3	-8	-50	-18
L Inferior Cerebellum	1	-14	-60	-28
L Lateral Cerebellum	1	-20	-46	-24

Table 3.3. Unisensory visual fMRI session activations: pointed > rounded, within cortical mask, FWE-corrected p < .05, cluster threshold 0 voxels; x, y, z: MNI coordinates for local maxima.

Region	**Cluster Size**	**x**	**y**	**z**
L Middle Cingulate Gyrus	4	-14	0	28
L Calcarine Sulcus (V1)	6	-14	-100	6
R Lingual Gyrus/Anterior Calcarine Sulcus	6	8	-62	4

Table 3.4. Unisensory visual fMRI session activations: rounded > pointed, within cortical mask, uncorrected p < .001, cluster threshold 4 voxels; x, y, z: MNI coordinates for local maxima.

To determine the extent to which congruency is determined via the distinction of rounded versus pointed in each unisensory modality, we contrasted pointed>rounded in the auditory session. This contrast of pointed > rounded in the auditory session within the cortical mask (FWE-corrected p < .05, cluster threshold 5 voxels) revealed multiple additional areas that may be important for the perception of sound symbolic crossmodal correspondences in the auditory modality (Table 3.5). As in the multisensory analyses, we see MFG and precentral gyrus in the pointed > rounded contrast for the auditory session. Similar to the pointed>rounded contrast in the visual session, we observe several left hemisphere cerebellar regions and the left FO, near Broca’s area.

In the auditory session, we also examined the reverse contrast of rounded > pointed within the cortical mask, and as for the equivalent contrast in the visual session, we found that this contrast was somewhat less robust (uncorrected p < .001, cluster threshold 4 voxels), but revealed several important areas (Table 3.6). Unlike in the multisensory analyses, areas such as the left superior frontal gyrus (SFG) and head of the hippocampus are present. As in the multisensory analyses, we see STG in the rounded > pointed contrast for the auditory session, but in the right hemisphere instead of the left.

Region	**Cluster Size**	**x**	**y**	**z**
L Middle Frontal Gyrus	1	-38	22	32
L Frontal Operculum (Broca's Area)	1	-32	6	4
L Precentral Gyrus	1	-50	6	22
	1	-56	2	38
L Inferior Precentral Sulcus (Premotor Cortex)	1	-26	-12	50
L Superior Cerebellum	1	-16	-46	-26
L Inferior Cerebellum	1	-16	-70	-22
L Medial Cerebellum	2	-18	-54	-26
	1	-18	-60	-26
	1	-8	-54	-26
R Medial Cerebellum	1	10	-48	-22

Table 3.5. Unisensory auditory fMRI session activations: pointed > rounded, within cortical mask, FWE-corrected p < .05, cluster threshold 0 voxels; x, y, z: MNI coordinates for local maxima.

Region	**Cluster Size**	**x**	**y**	**z**
L Superior Frontal Gyrus	4	-10	24	52
R Planum Temporale/Superior Temporal Gyrus	7	58	-2	2
R Head of Hippocampus	4	24	-38	14

Table 3.6. Unisensory auditory fMRI session activations: rounded > pointed, within cortical mask, uncorrected p < .001, cluster threshold 4 voxels; x, y, z: MNI coordinates for local maxima.

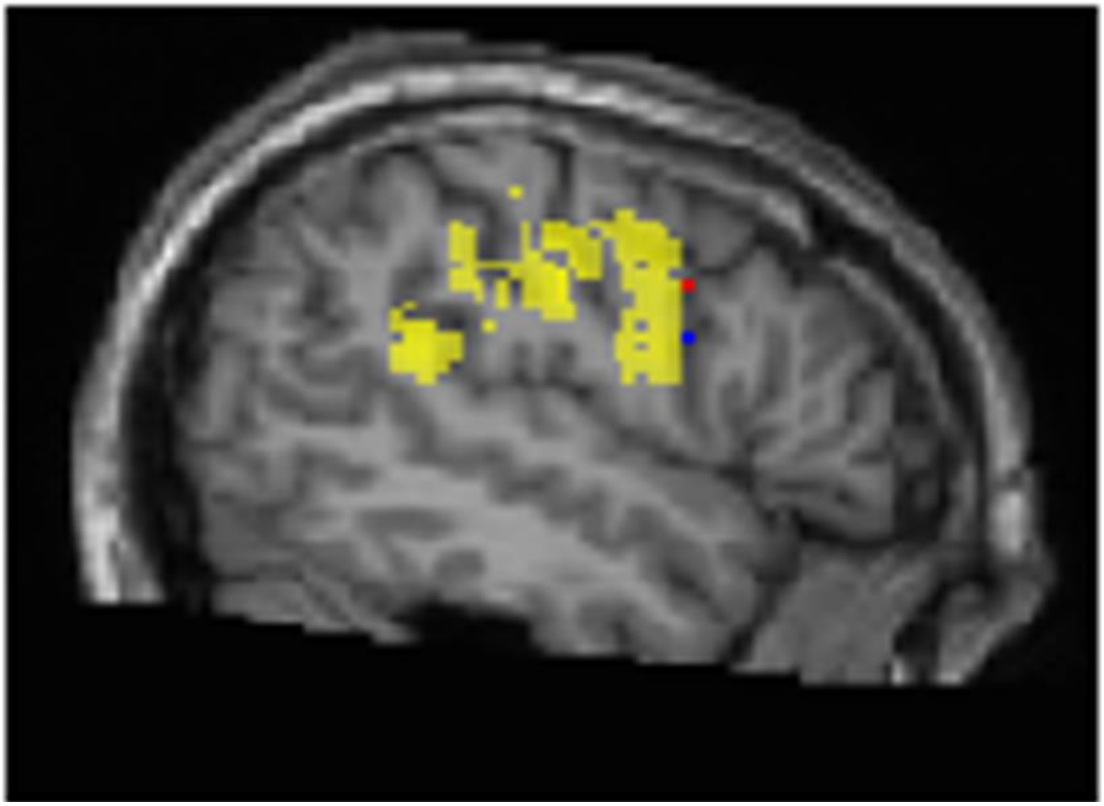


Figure 3.6. View of the left precentral gyrus-postcentral gyrus cluster (cluster size 1601 voxels) at the second level group analysis (n = 24) in yellow on the multisensory session with a univariate contrast of congruent > incongruent (FWE $p < 0.05$). Overlaid on this view is the contrast of pointed>rounded for visual (FWE $p < 0.05$) and auditory sessions (FWE $p < 0.05$), respectively. View is a sagittal view in which left is posterior and right is anterior. Shown here are the left precentral gyrus (MNI -40,-12, 30) in red in the visual session and the left inferior precentral gyrus in blue (MNI -26, -12, 50) in the auditory session. These areas are adjacent to the left precentral gyrus-postcentral gyrus cluster from the multisensory session.

There are some areas in which the pointed>rounded contrast for the unisensory sessions is adjacent or overlapping with the multisensory congruent>incongruent contrast. For example in Figure 3.6, the view of the left precentral gyrus (MNI -40,-12, 30) in the visual session in red and the left inferior precentral gyrus (MNI -26, -12, 50) in the auditory session in blue are shown. These areas are adjacent to the left precentral gyrus-postcentral gyrus cluster from the multisensory session.

We then sought to investigate the potential connection between the pseudoword and shape judgments of roundedness and pointedness and the neural response to these stimuli. In order to investigate this objective, we conducted a correlational item analysis in which each stimulus, instead of each participant, was modeled in the random effects analysis at the second level, and the mean rating for each stimulus was a covariate. Neither of these analyses survived FWE-correction, but the results for the visual (Table 3.7) and auditory (Table 3.8) correlational analyses hints that the roundedness/pointedness ratings for the stimuli were correlated with stimulus>baseline activation in multiple areas (uncorrected $p<0.001$, cluster threshold of 4 voxels).

For the visual correlational analyses, similar to the pointed>rounded contrast in the visual and auditory sessions, we observe the FO, but in the right hemisphere instead of the left. We also observe the right MTG in the visual correlational analysis, similar to the left MTG in the multisensory congruent>incongruent contrast. As in the multisensory analyses and the visual session contrast of pointed>rounded, angular gyrus was evident in the right hemisphere rather than the left. Also like the multisensory analyses, the visual correlational analysis revealed several areas in the right SMG.

For the auditory correlational analyses, we observe putamen, which has not previously appeared in our univariate analyses. As observed in the visual pointed>rounded contrast, we see the left insula in the auditory correlational analysis, and for the multisensory congruent>incongruent contrast, insula appeared in the right hemisphere. Interestingly, activation in left SOG is observed in the auditory correlational analysis and not the visual correlational analysis, and activation in right SOG was observed in the multisensory analyses.

Region	Cluster Size	x	y	z
L Supramarginal Gyrus	4	-42	-34	38
R Opercular part of Inferior Frontal Gyrus	10	54	8	14
R Posterior Middle Temporal Gyrus	9	56	-58	-4
R Supramarginal Gyrus	7	48	-26	48
	6	56	-18	36
	5	58	-26	42
	4	52	-26	38
R Superior Parietal Lobule /Angular Gyrus	4	40	-46	54

Table 3.7. Unisensory visual fMRI session activations: correlations with perceptual ratings of each stimulus, within cortical mask, uncorrected p < .001, cluster threshold 4 voxels; x, y, z: MNI coordinates for local maxima.

Region	Cluster Size	x	y	z
L Superior Occipital Gyrus	4	-16	-94	16
L Putamen/Post Insula	4	-30	-16	-2
L Putamen	6	-22	-2	-10

Table 3.8. Unisensory auditory fMRI session activations: correlations with perceptual ratings of each stimulus, within cortical mask, uncorrected p < .001, cluster threshold 4 voxels; x, y, z: MNI coordinates for local maxima.

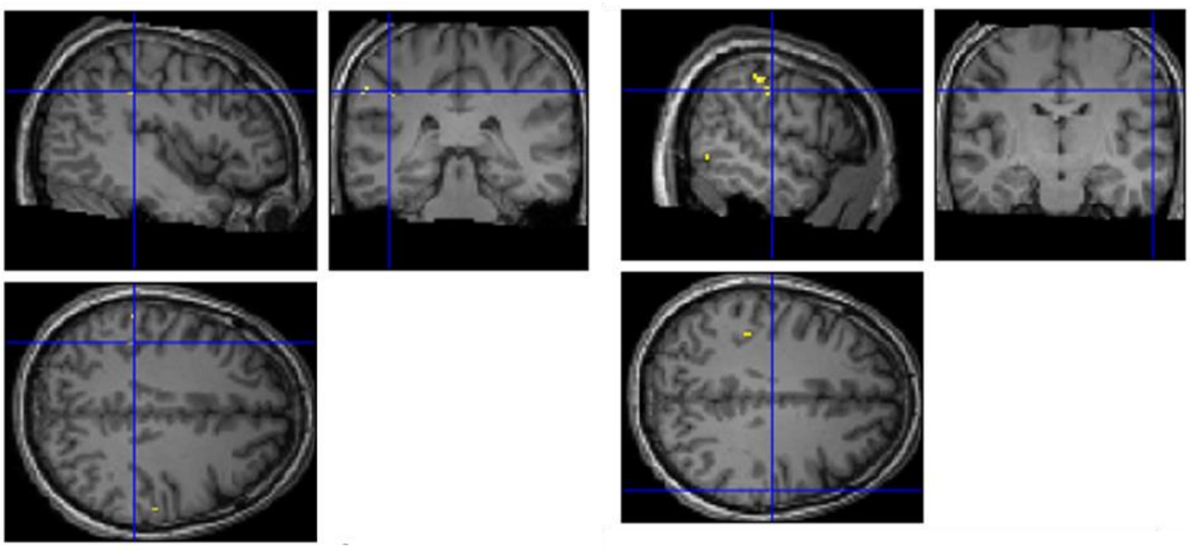


Figure 3.7. On the left is the view of the left supramarginal gyrus cluster (cluster size 4 voxels; MNI -42, -34, 38) in the unisensory visual session correlated with the visual perceptual ratings (uncorrected p<0.001) in yellow. On the right is the view of the right supramarginal gyrus cluster (cluster size 7 voxels; MNI 48, -26, 48) in the unisensory visual session correlated with the visual perceptual ratings (uncorrected p<0.001) in yellow. Panels include a sagittal (top left), coronal (top right), and transverse (bottom left) view. Within the transverse panel, right is anterior, left is posterior. Within the coronal panel, left is left hemisphere, and right is right hemisphere.

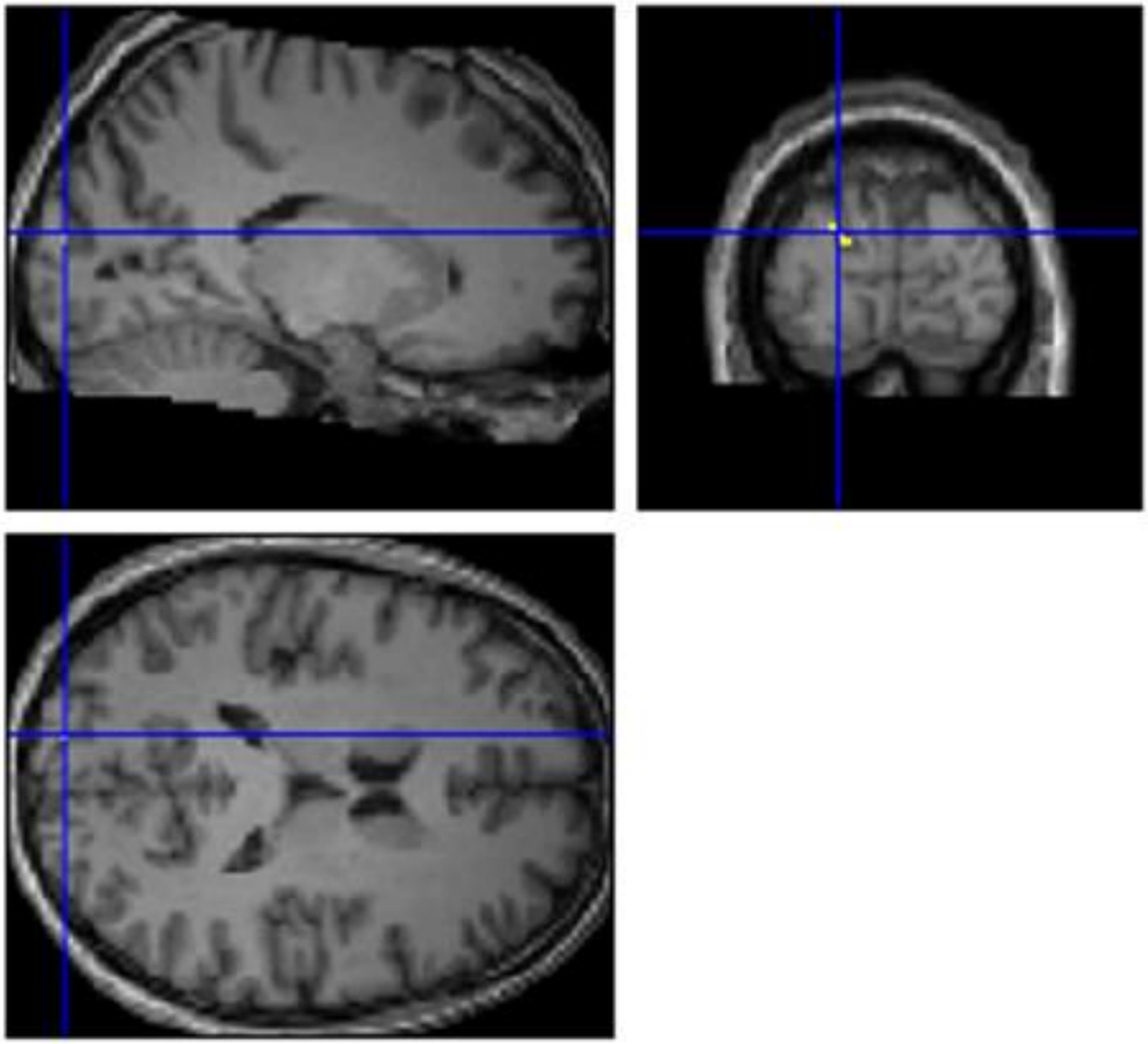


Figure 3.8. A view of the left superior occipital gyrus cluster (cluster size 4 voxels; MNI -16, -94, 16) in the unisensory auditory session correlated with the auditory perceptual ratings (uncorrected p<0.001) in yellow. Panels include a sagittal (top left), coronal (top right), and transverse (bottom left) view. Within the transverse panel, right is anterior, left is posterior. Within the coronal panel, left is left hemisphere, and right is right hemisphere.

After investigating each of the univariate contrasts, we returned to the left precentral gyrus-postcentral gyrus cerebral cortical region from the multisensory congruent>incongruent contrast (local maximum -48,-16,30; cluster size 1601) to compare the differential activation in that area for both the multisensory and unisensory sessions. Figure 3.9 a. illustrates the positive contrast estimate for the multisensory congruent>incongruent contrast, from which this cerebral cortical region was derived. Figure 3.9 b. illustrates the contrast estimate for the multisensory incongruent>congruent contrast, and this contrast is zero within the 90% confidence intervals. Figure 3.9 c. illustrates the contrast for the visual pointed>rounded contrast, and this contrast was positive indicating that the cerebral cortical region derived from the multisensory congruent>incongruent contrast also showed selectivity for pointed visual stimuli. Figure 3.9 d. illustrates the contrast estimate for the visual rounded>pointed contrast, and this contrast was zero, indicating that this cerebral cortical region does not show selectivity, via differential estimates of the BOLD signal, for rounded over pointed visual stimuli. Figure 3.9 e. illustrates the contrast estimate for the auditory pointed>rounded contrast, and this contrast was positive indicating that the cerebral cortical region derived from the multisensory congruent>incongruent contrast also showed selectivity for auditory stimuli that were rated as pointed. Figure 3.9 f. illustrates the contrast estimate for the auditory rounded>pointed contrast, which is zero and indicates that this cerebral cortical region did not show selectivity for pseudoword stimuli that were rated as rounded over pointed. We completed this analysis to probe how this cerebral cortical area responds under conditions that were multisensory and to compare this activity to conditions that were unisensory.

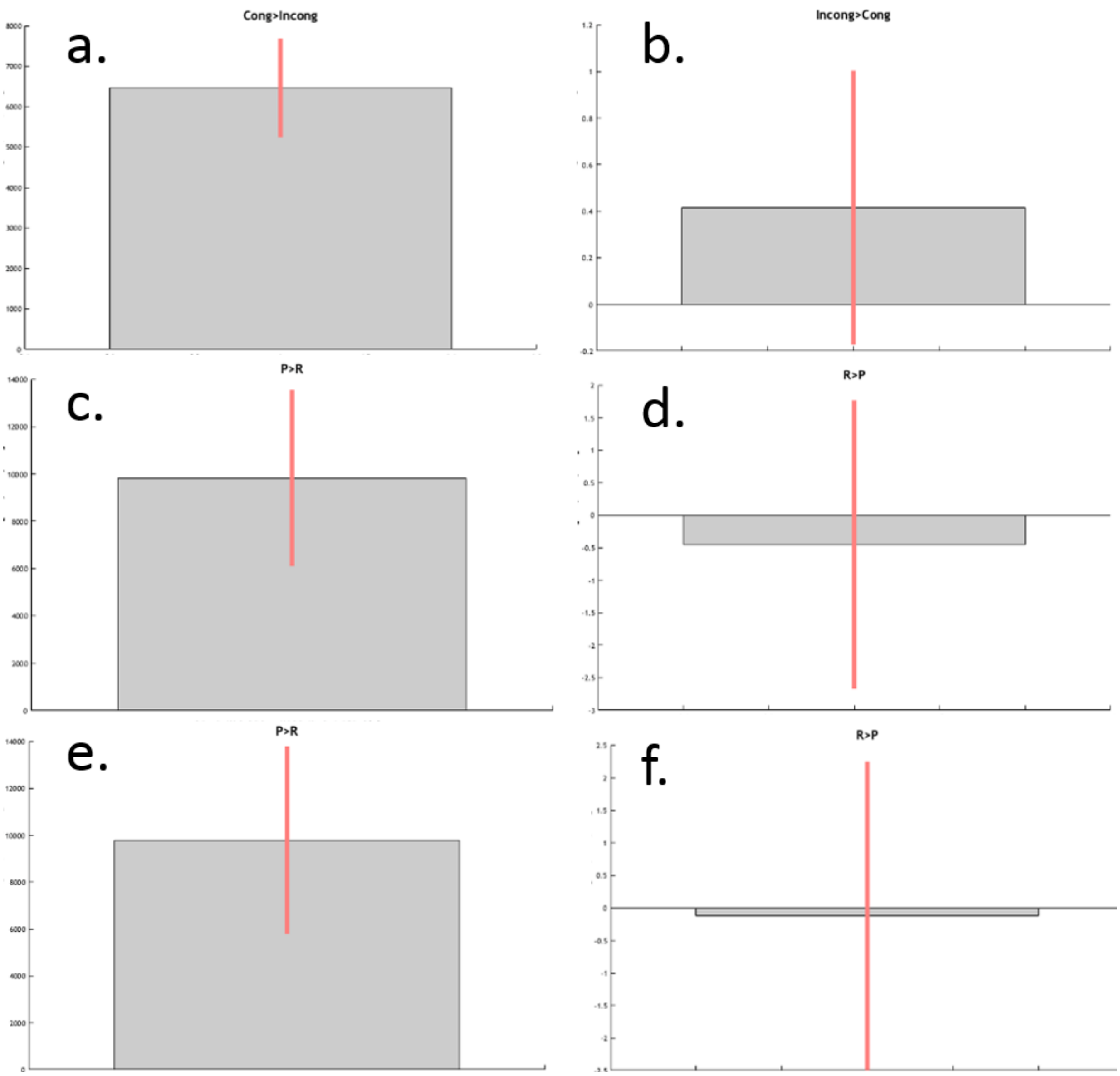


Figure 3.9. Contrast beta value estimates with 90% confidence intervals (red) within the multisensory left precentral gyrus-postcentral gyrus (local maximum -48,-16,30 x,y,z: MNI coordinates, cluster size 1601 voxels). A. illustrates the positive contrast estimate for the multisensory congruent>incongruent contrast, from which this cerebral cortical region was derived. B. illustrates the contrast estimate for the multisensory incongruent>congruent contrast, which is statistically not greater than zero. C. illustrates the positive contrast for the visual pointed>rounded contrast. D. illustrates the contrast estimate for the visual rounded>pointed contrast, which is statistically not greater than zero. E. illustrates the positive contrast estimate for the auditory pointed>rounded contrast. F. illustrates the contrast estimate for the auditory rounded>pointed contrast, which is statistically not greater than zero.

In sum, the univariate contrasts of pointed > rounded and rounded > pointed while participants perceived the unisensory components of sound-symbolic CCs in the scanner revealed multiple areas, primarily in the left hemisphere but also in the right that may contribute to the processing of sound-symbolic CCs. Interestingly, the correlational analyses with the roundedness/pointedness ratings revealed SMG in the right hemisphere for the visual session, an area that mirrors the left SMG that is important for language, and the visual cortical area L SOG for the auditory session. Further, the left precentral gyrus-postcentral cerebral cortical region from the multisensory congruent>incongruent contrast illustrates a sensitivity to both congruent over incongruent auditory-visual sound symbolic CCs but also a sensitivity to pointed stimuli over rounded stimuli in both the visual and in the auditory modalities. Taken together with the univariate results of the multisensory fMRI task, these regions may be important for the processing of sound-symbolic CCs, for speech or language processing, for sensory processing, and even auditory-visual multisensory integration more generally.

3.4.2.2 Multivariate Analyses with Unisensory Visual and Auditory Task Data

We used the univariate analysis of the multisensory session to identify multiple candidate regions of interest (ROIs) for multivariate analyses. We then performed a multivariate analysis called representational similarity analysis (RSA) on those ROIs using unisensory data to determine how the potential multisensory regions that we previously identified may be important for sound-symbolic CCs in a unisensory context. For each ROI and each participant, we performed RSA for the t-statistic of the BOLD response comparing each stimulus to baseline. Once we obtained one RDM for each ROI for each participant for the data obtained while they were viewing shapes, we compared that RDM to the RDM from their rounded/pointed ratings of those same visual shapes using a second order Spearman correlation. We used this approach to investigate correlations between perceptual ratings and neural activation for both the auditory and visual sessions and to investigate correlations crossmodally. We performed the same procedure for the auditory stimuli for a total of 912 neural RDMs across modality, participant, and ROI. To summarize these data, Figure 3.10 display the regions of interest from Table 3.2 on the horizontal axis, and the vertical axis displays the average second order Spearman correlations between visual neural RDMs and visual perceptual RDMs and the standard error of those correlations across participants (as detailed in Methods Section 3.3.8) . In Figures 3.10, 3.11, and 3.12, bars represent standard errors for the second order Spearman correlation in each ROI. However, none of those correlations survived statistical significance.

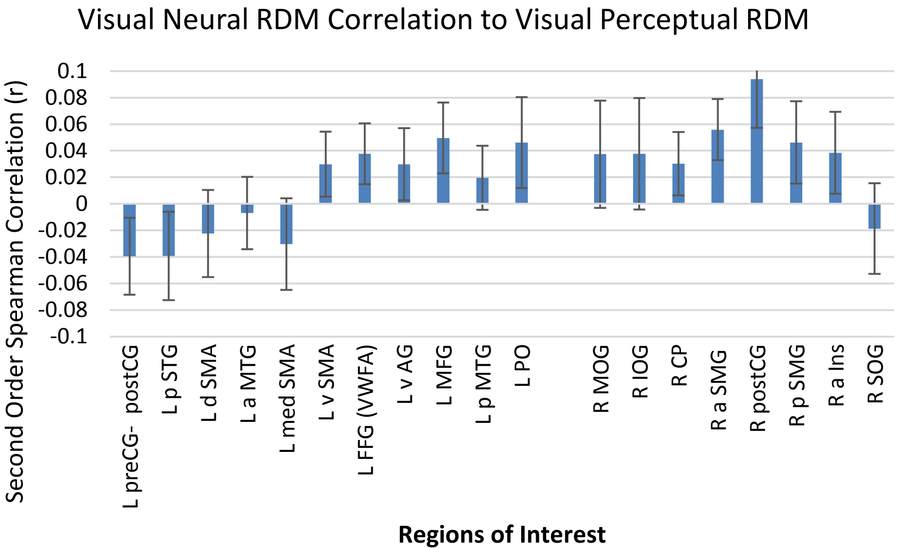


Figure 3.10. The RDMs for the neural response to shapes for the ROIs from the multisensory univariate analysis compared to the RDMs for the judgments of these shapes did not reveal any significant correlations across participants.

To determine the functional significance of our multisensory ROIs during the perception of auditory pseudowords alone, we also obtained one RDM for each ROI for each participant for the data obtained while they were listening to pseudowords. We then compared that RDM to the RDM from their roundedness/pointedness ratings of those same auditory pseudowords using a second order Spearman correlation (as detailed in Methods Section 3.3.8). To summarize these data, Figure 3.11 displays the regions of interest from Table 3.2 on the horizontal axis, and the vertical axis displays the average second order Spearman correlations between auditory neural RDMs and auditory perceptual RDMs and the standard error of those correlations across participants. These data are more variable than those for the visual RDMs, in part because there was more individual variation (See Chapter 4, Section 4.4.1) across participants in their perceptual ratings of the pseudowords. However, none of those correlations reached statistical significance.

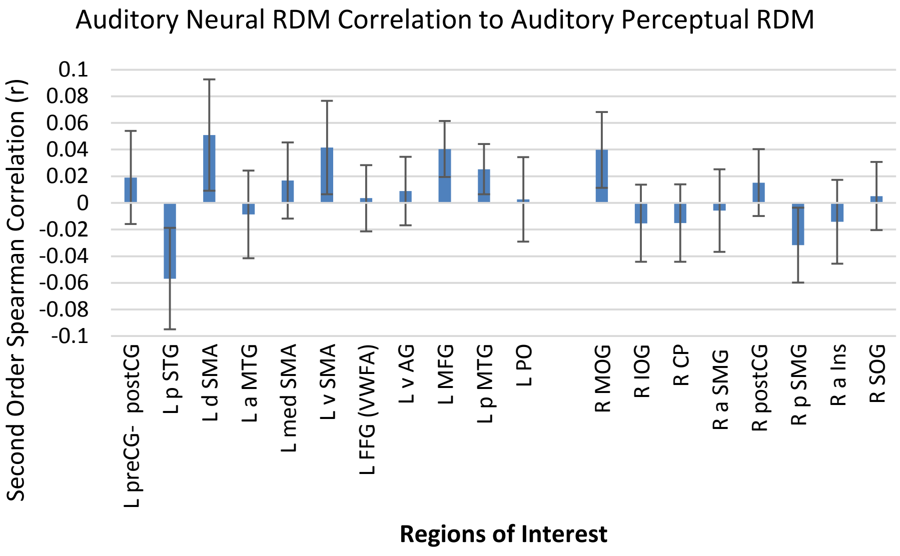


Figure 3.11. The RDMs for the neural response to pseudowords for the ROIs from the multisensory univariate analysis compared to the RDMs of the judgments of these pseudowords did not reveal any significant correlations across participants.

We next sought to determine whether any of the ROIs from the multisensory session were crossmodally similar in a visual unisensory task and an auditory unisensory task using a range of stimuli that vary from rounded to pointed sound symbolically (List et al., BioRxiv; McCormick et al., 2015). For each ROI and each participant, we then compared each neural RDM from the visual shape fMRI task to the RDM from the auditory pseudoword fMRI task using a second order Spearman correlation. To summarize these data, Figure 3.12 displays the regions of interest from Table 3.2 on the horizontal axis, and the vertical axis displays the average second order Spearman correlations between visual neural RDMs and auditory neural RDMs and the standard error of those correlations across participants. However, none of those correlations reached statistical significance.

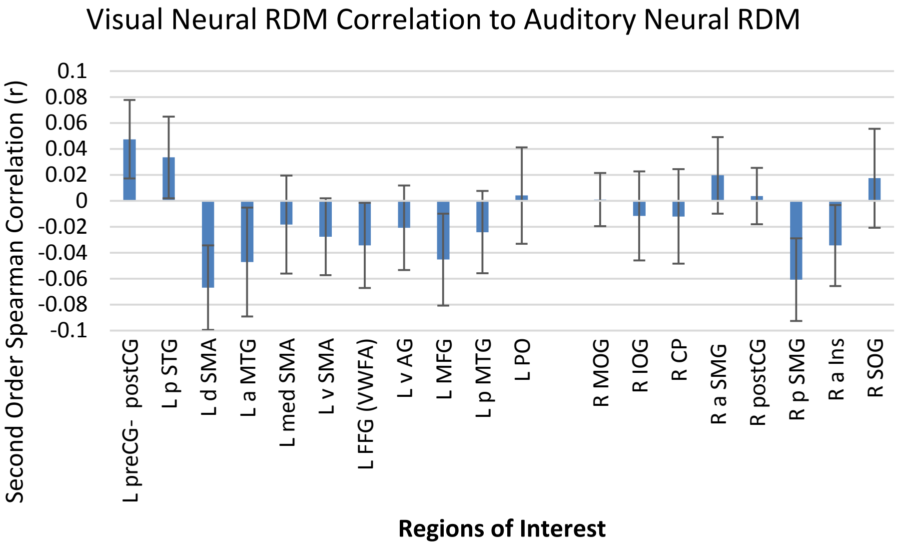


Figure 3.12. The RDMs for the neural response to shapes for the ROIs from the multisensory univariate analysis compared to the RDMs for the neural response to pseudowords did not reveal any significant correlations across participants.

In sum, our multivariate analysis using RSA to compare perceptual ratings to unisensory perception of both auditory pseudoword and visual shapes revealed that for this particular paradigm, RSA was not the ideal method to recapitulate the effects that are observable in the univariate analyses, indicating that there may be other multivariate methods, such as searchlight approaches, that would be able to add nuance, but these analyses are beyond the scope of this dissertation.

3.5 Discussion

While others have examined the neural underpinnings of sound symbolism with standard univariate methods (Revill et al., 2014; McCormick et al., 2018), The present study is the first to systematically examine the multisensory and unisensory neural correlates of sound symbolism in the sound-shape domain using both univariate and multivariate approaches. We are the first to employ both shapes and pseudowords in a multisensory and unisensory context to investigate the visual, auditory, and multisensory brain regions that are important for sound-symbolic crossmodal correspondences.

We here replicated the behavioral findings from Chapter 2 using a subset of the stimulus set (List et al., BioRxiv), illustrating the wide range of stimuli over which sound symbolic crossmodal correspondences can occur and that a subset of these stimuli can replicate our previous findings. This advantage allowed us to explore a spectrum of rounded to pointed shapes and pseudowords. Our study is also the first to investigate the similarities in perceptual judgments of roundedness/pointedness for shapes compared to pseudowords and to further use these similarities to determine how visual and auditory neural capabilities are important for sound-symbolic CCs both within and across modalities. Despite using different properties to make sound-symbolic judgments, raters were able to use a common scale of roundedness/pointedness across the two different modalities, as indicated by the similarity in the RDMs for shape versus pseudoword judgments. This is not surprising, given that participants were asked to use the same scales in both modalities, but the fact that different groups of participants, using two opposing scales in each modality, produced ratings whose (90x90) dissimilarity matrices were highly correlated across modalities, is evidence that the sound-symbolic CC studied here is robust. The strong crossmodal correlation provides a firm basis for further exploration of the principal aim of the present study, characterizing stimulus properties that may underpin sound symbolism.

Previous fMRI studies of sound-symbolic crossmodal correspondences have investigated the neural locus of sound symbolic synonym-antonym pairs (Revill et al., 2014) or of a single set of sound symbolic stimuli (McCormick et al., BioRxiv). Other ERP studies of sound symbolic crossmodal correspondences have investigated a range of sound-symbolic words and pseudowords (Kovic et al., 2010; Sučević et al., 2015; Lockwood, Hagoord, & Dingemanse, 2016), which may implicate different neural mechanisms depending on many factors. These factors include but are not limited to the length of the auditory stimuli, the acoustic or phonetic properties of the stimuli (See Chapter 2), and whether or not the stimuli are real words or pseudowords.

The results of Peiffer-Smadja & Cohen, 2019 are another fMRI investigation into the neural correlates of sound symbolic crossmodal correspondences. As with our study, Peiffer-Smadja & Cohen, 2019 investigated a range of shapes (36 total stimuli) and pseudowords (48 total stimuli) in both unimodal and bimodal conditions. In contrast to our study, Peiffer-Smadja & Cohen found activations in the left and right dorsolateral prefrontal cortices for the contrast of incongruent>congruent (left: MNI -42, 44, 22; right: MNI 34, 56, 28) where our study did not find activations that survived FWE correction. However, our study did reveal left middle frontal gyrus (MNI -32, 40, 22 cluster size 7) in the multisensory contrast of congruent>incongruent (Table 3.2). In addition, for the contrast of pointed>rounded, in the visual session we obtained frontal areas in left inferior frontal gyrus (MNI -36, 8, 30 cluster size 1) and in the auditory session, left middle frontal gyrus (-38, 22, 32 cluster size 1), indicating that frontal regions may be important for the distinction of matching or mismatching via the determination of rounded or pointed in two sensory modalities. The left middle frontal gyrus has been shown in Peiffer-Smadja & Cohen (2019) and in McCormick et al. (BioRxiv) to be preferentially activated for conditions of incongruency (See also, Hein et al., 2007), which may raise the question of why this region appears in our studies preferring a condition of congruency. The middle frontal gyrus not only appears consistently in analyses involving congruency, but also is part of the multiple demand attentional system that is not preferential to a specific sensory modality (Duncan, 2013). One possibility is that the pairs of stimuli that match are more environmentally salient because consistency across sensory modalities suggests the resulting signal is conferring accurate information to the brain (Fetsch et al., 2013). However, certain behavioral tasks, perhaps including those of Peiffer-Smadja & Cohen (2019) or McCormick et al. (BioRxiv) the attempt to integrate induces greater BOLD signal for mismatching than matching pairs of stimuli because the brain has failed to integrate (Noppeney et al., 2012).

For Peiffer-Smadja & Cohen (2019), the contrast of kiki>bouba did not reveal significant activations, but in our results, the pointed>rounded or kiki>bouba contrast revealed more significant (FWE-corrected) activations than the opposite contrast of bouba>kiki or rounded>pointed (Table 3.5 & 3.6). Peiffer-Smadja & Cohen (2019) were combining unimodal and multimodal stimuli in the same contrast for their bouba>kiki or rounded>pointed contrasts, which may have been done to increase power of the results but is in opposition to the basic principle of MSI known as supra-additivity (See Chapter 1). Supra-additivity indicates that a multisensory pair of stimuli yields a response that is greater than the simple summation of the response to a single stimulus in either modality. This principle in terms of fMRI is not always adhered to in the same manner as the original electrophysiological experiments (Sten & Meredith, 1993), but is often taken to indicate that the neural mechanism, which can include the size of the response or even the area from which the response is elicited, could be different in the unisensory and in the multisensory condition (Murray et al., 2018; Noppeney, 2012). Therefore, combining a unisensory and a multisensory condition for the same stimuli may be a statistical equivalent of the combination of apples and oranges. Indeed for Peiffer-Smadja & Cohen (2019), the bouba>kiki contrast was more robust in blocks for which the paired shape was matching, or congruent, than in blocks for which the paired shape was mismatching, confirming that a multisensory condition can be, and often is, represented differently than a unisensory condition.

Our study consisted of separate sessions for each auditory unisensory, visual unisensory, and a third multisensory session to avoid spillover in the time course of the BOLD signal and to minimize the possibility of inflating participants' assumptions about the relationships among the stimuli. Peiffer-Smadja & Cohen's (2019) attempts to delineate the neural correlates of sound symbolic crossmodal correspondences are rigorous and sound. However, we suspect that our deference to the original principles of multisensory integration have contributed to the differences in our respective results, including the additional power afforded to our studies by having separate sessions for unimodal and bimodal conditions and the fact that we analyzed stimuli presented in a unimodal and in a bimodal task as separate conditions.

Both of the unisensory sessions were done separately and with the auditory session first to minimize participants' assumptions about how the auditory pseudowords might be related to the visual shapes. Given that fact, it was surprising to us that during the correlational results in the visual session, areas that are often implicated in auditory phonological processing such as bilateral supramarginal gyrus (Hartwigsen et al., 2010) were correlated with average visual perceptual ratings. The converse was also true in that during the correlational results, in the auditory session, an area that is often implicated in visual processing of letters and shapes, the left superior occipital gyrus (Fullbright et al., 2003) was correlated with average auditory perceptual ratings. From these results, one can conclude that even in the absence of stimuli from one sensory modality (e.g. vision), individuals are engaging areas that are primarily associated with processing that modality (vision) while experiencing the stimuli from the other modality (e.g. hearing auditory pseudowords). Further, this activation was correlated with the perceptual ratings for the modality that is being experienced (e.g. auditory pseudowords). Perhaps this finding indicates that, particularly for the perceptual ratings of pseudowords, visual imagery plays an important role. There may be phonetic elements that co-occur with certain environmental stimuli, just as we investigated in Chapter 2 that certain acoustic properties tend to couple with roundedness and pointedness (Nielsen & Rendall, 2011; Ozturk et al., 2013; Fort et al., 2015; See also Sidhu & Pexman, 2017 for review). Future directions may investigate a correlation between individuals' imagery preference and the degree to which the left superior occipital gyrus is correlated with perceptual ratings of pseudowords, similar to the methods of McCormick et al. (BioRxiv) in correlating activation magnitudes of the incongruent>congruent contrast and individuals' preference for object over spatial imagery.

The one individual who decided that matching pairs of stimuli were those that were canonically incongruent raised some interesting questions about the relationship between multisensory perception and multisensory integration. One possibility is that this individual does not integrate sensory inputs in the same way as other participants. A more nuanced discussion of the individual differences of the group sensitivity to sound symbolic crossmodal correspondences is present in Chapter 4. Noppeney (2012) has argued that the brain may unsuccessfully attempt to integrate incongruent multisensory inputs, which in some cases can lead brain regions to respond more to incongruent conditions than congruent ones. Whether or not this failure to integrate in the brain is dependent on what is perceived as "matching" is an open question. Although the contrast for the group of incongruent>congruent yielded no results that survived FWE correction, perhaps the incongruent>congruent for the oddball individual would yield greater activations than the same contrast for another individual in the group, either because of a failure to integrate or because that participant's unique perception led to integration more similar to the congruent>incongruent contrast in others. Future directions may compare the neural correlates for a group of individuals who indicate the opposite sensitivity to sound symbolic crossmodal correspondences (i.e. "matching" pairs of stimuli are incongruent) with the neural correlates for a group of individuals who show canonical sensitivity to these associations (i.e. "matching" pairs of stimuli are congruent). If individuals' integration of multisensory inputs is dependent on the perception of “matching” stimuli, perhaps a group who decided “matching” pairs of stimuli are those that are canonically incongruent would have a neural response for a congruent>incongruent contrast that was similar to the contrast of incongruent>congruent for the typical group. The relationship between the perception of multisensory matching and the neural mechanisms of the multisensory integration is intriguing from the standpoint of understanding the link between perception and neural activation.

3.6 Conclusion

Although sound-symbolic crossmodal correspondences have been studied for almost a century, few have systematically examined the neural regions of both the auditory and visual dimensions that elicit sound-symbolic CCs (McCormick et al., BioRxiv; Peiffer-Smadja & Cohen, 2019). Here, we used functional magnetic resonance imaging (fMRI) and both univariate correlational analyses and representational similarity analysis to examine the relationships between perceptual ratings that varied on dimensions of roundedness and pointedness, for a range of auditory pseudowords and visual shapes and the neural response to these stimuli. The univariate contrast of congruent > incongruent while participants perceived sound-symbolic CCs in the scanner revealed multiple areas, primarily in the left hemisphere but also in the right, that may be important for the processing of sound-symbolic CCs, for speech or language processing, for sensory processing, and even auditory-visual multisensory integration more generally. Our multivariate analysis using RSA to compare perceptual ratings to unisensory perception of both auditory pseudoword and visual shapes revealed no regions that survived statistical testing, but this finding does not preclude the hypothesis that there may be cerebral cortical regions that have a modulatory role in unisensory perception of crossmodal correspondences. The fact that individuals can rate pseudowords on a visual property in the absence of visual shape indicates that there may be unisensory modulation of the perception of sound symbolic crossmodal correspondences, though this phenomenon is typically considered multisensory. This idea is further supported by the correlational results in which we found a visual area, left SOG to be activated in the unisensory auditory session as correlated with pseudoword ratings. Further, the correlational results indicated an auditory area, bilateral SMG, to be activated in the unisensory visual session as correlated with shape ratings. These modulatory functions in the respective unisensory domains could then contribute to their multisensory role in the perception of language, particularly in the perception of sound-symbolic CCs.

These findings suggest that the perception of sound-symbolic CCs is complex and multimodal but that canonically multisensory regions (e.g. the posterior superior temporal gyrus) may not explicitly regulate these associations in a unisensory context alone. This research provides insights into the fundamental nature of CCs and how they might evoke specific interpretations of physical meaning in natural language.

Chapter 4

Synesthesia & Individual Differences in Perception of Sound-symbolic Crossmodal Correspondences

4.1 Abstract

One of the fundamental properties of language is the assumed arbitrariness of the relationship between sound and meaning, but sound symbolism refers to non-arbitrary mapping between the sound of a word and its meaning. One way that sound symbolism has been studied is through crossmodal correspondences (CCs) between auditory pseudowords (e.g. ‘loh-moh’) and visual shapes (e.g. a blob). Sound-symbolic CCs may implicate the process of auditory-visual multisensory integration, a process by which the input from multiple sensory modalities is combined into a single coherent consciousness. Although sound-symbolic crossmodal correspondences are consistent across language background, cultural background, and children pre-reading age, one of the fundamental principles of MSI is that it is context-dependent. Here, we expand on experiments presented in Chapter 3 and explore this purported paradox between a consistent phenomenon that is supported by a highly flexible neural process. We examine individual differences in twenty-four typical individuals and one grapheme color synesthete as an example of extreme perceptual difference. We have here demonstrated that even within the same experimental context, multiple individuals can perceive and respond quite differently to sound-symbolic CCs behaviorally. These differences are evident neurally in the qualitative similarities and differences evident in one synesthete, one example participant who performed well in the behavioral assays of sound-symbolic CCs, and the second level group analysis (n = 24) during a multisensory task. These findings may support the theory that synesthetes’ increased sensations in auditory and visual modalities extends to increased sensitivity to sound-symbolic CCs. This research provides insights into the fundamental nature of CCs and how they function within the context-dependent nature of MSI.

4.2 Introduction

We have earlier established the consistent nature of sound-symbolic crossmodal correspondences (CCs; List et al., BioRxiv; See Chapter 1). Sound-symbolic CCs involve associations between seemingly unrelated auditory and visual features (Spence, 2011) but they are also highly consistent with individuals across different language backgrounds (Blasi et al., 2016), different cultural backgrounds (Chen et al., 2016; Kantartzis et al., 2011), and with children pre-reading age (Imai et al., 2015; Ozturk et al., 2013; Tzeng et al., 2017). We have also established that the neural processing of sound-symbolic CCs is multisensory in nature in that it involves both auditory and visual cortices (See Chapter 2; See Chapter 3). Here we investigate through the lens of sound-symbolic CCs the individual variation in perception in typical individuals and a grapheme-color synesthete and how that variation leads to consistent patterns at the group level.

4.2.1 Individual variation in multisensory integration

Since multisensory integration (MSI) is an adaptive process that is context-dependent (See Chapter 1, Section 1.2), one may necessarily question to what extent individual differences in perception can be explained by variation in these neuronal mechanisms of MSI and vice versa. Each individual's perceptual experiences vary, but how these variations may shape the characteristics of their neural or behavioral response to multisensory stimuli is not known. According to a computational model of MSI, the Bayesian estimator, the priors, or measurements of unknown information, are different for each individual taking different measurements of the same stimuli, based on their sensory receptors, their attentive state, their genetics, or their past experiences, and MSI is subsequently influenced by that individual variation (Welch & Warren, 1980).

For example, increased sensitivity to auditory-visual synchrony has been found from perceptual and from musical training (Lee & Noppeney, 2011; Powers et al., 2009). Powers et al., 2009 used a task in which participants judged auditory-visual simultaneity with feedback, and found that this feedback fine-tuned participants' temporal sensitivity through multisensory learning. This study highlights the flexibility of temporal multisensory processing and further suggests that experience may calibrate the accuracy of temporal predictions through crossmodal cues, perhaps through increased neural synchrony across modality. In addition, the default modality for sensory dominance or bias has been shown to vary across individuals (Romei et al., 2013) and may also be sensitive to training (Sandhu & Dyson, 2012). The effects of individual experience in MSI can develop over a lifetime or can develop rapidly due to brief training, as have been shown in recalibration experiments in the domains of time, space, and speech content (Hillock-Dunn & Wallace, 2012; Stevenson et al., 2018; Van der Burg et al., 2013; Vroomen & Baart, 2012).

If neural variation in auditory-visual MSI may be altered due to experience, then individual variation in behavioral measures of sound-symbolic crossmodal correspondences (CCs) would not be surprising. In fact, earlier chapters have discussed some of the potential differences in sound-symbolic CCs across cultures (Chen et al. 2016). However, if sound-symbolic associations are largely consistent across languages (Blasi et al., 2016) and these consistencies lead to advantages to language processing in adults and children (Brand et al., 2018; Gasser et al., 2004; Nygaard et al., 2009; See Chapter 2), then the question arises how so much individual variation in MSI allows for sound-symbolic CCs in the first place. One possible explanation is proposed by Ramachandran and Hubbard (2001) who developed a theory of the evolution of language based on studying synesthetes and on evidence from crossmodal correspondences in other domains, such as pitch and brightness (Marks, 1987; Melara & Marks, 1990a; Melara & Marks, 1990b; Melara & O'Brian, 1987). For example, high-pitched tones are consistently judged to correspond with bright hues, and these associations were once referred to as synesthetic, indicating a joining of the senses, because they reflect a presumed connection among attributes from different sensory modalities, though the term synesthetic associations in reference to typical individuals is no longer preferred in the field (Melara & O'Brian, 1987). Ramachandran and Hubbard use these correspondences, which they refer to as synesthetic correspondences, to propose that language arose from sensory-to-motor synesthesia, such that certain lip and tongue movements represented in motor areas may be systematically mapped to certain phonemic and acoustic perceptions mapped in auditory areas of the brain. Further, perhaps synesthesia is an extreme on a spectrum of perceptual individual variation, which can be measured through differences in sound-symbolic CCs (Lacey et al., 2016).

4.2.2 Synesthesia as an extreme individual perceptual difference

4.2.2.1 Describing synesthesia

Synesthesia is a phenomenon in which a person responds to a primary stimulus in one modality, or the "inducer", with a secondary, synesthetic experience in a separate modality, called the "concurrent" (Eagleman et al., 2007). Examples of synesthetic experiences include hearing music and concurrently seeing colors or seeing movement and concurrently feeling textures. Synesthesia can also occur within one modality. While most people reading this text see a series of letters in the color black, grapheme-color synesthetes either see or strongly and consistently associate different colors with letters and numbers, such that this page appears to have an explosion of colors. Grapheme-color synesthesia is the most common type of synesthesia, occurring in the general population at a rate of 1-2% of individuals (Simner et al., 2012), but there is a wide range of synesthetic experiences. One synesthete insists that the taste of spicy chicken is pointy (Cytowic, 2003), and another synesthete personifies ordinal linguistic sequences such that they report that the number "8 is a very dignified lady, who…is the wife of 9 [who is] self-centered, maniacal, grumpy, endlessly reproaching his wife for one thing or another" (Flournoy, 1893, pp. 219-220; Simner & Holenstein, 2007).

Given the varied and unusual nature of synesthetic experiences, perhaps it is unsurprising that synesthesia was not part of mainstream science for several decades, but cases of synesthesia have been reported for over 200 years (Sachs, 1812). There are records of as many as 61 types of synesthesia, and color is the most common secondary experience (Day, 2005).

Reports on the incidence of synesthesia in the general population vary widely from as high as 4% (Simner et al., 2006), from 1-2% (Simner et al., 2012), or even as low as 0.05% (Baron-Cohen et al., 1996). A person with synesthetic relatives is more likely to develop synesthesia than someone without synesthetes in their family (Galton, 1880; Baron-Cohen et al., 1996). Although the findings were heterogeneous, Tilot et al. (2018) have implicated a genetic component to synesthesia, especially in relation to axonogenesis and neuronal development. Those that are synesthetic may be more likely to be involved in the arts (Rich et al., 2005). While early reports suggested that synesthesia is also more common in individuals who are left-handed (Cytowic, 1989), more recent studies have not found an association between left-handedness and the development of synesthesia (Rich et al., 2005).

Comparing the experiences among synesthetic individuals, researchers have found that synesthetic associations can be characterized by three traits: arbitrariness or randomness of the development of the synesthetic associations, automaticity or the inability to suppress the associations, and idiosyncrasy, meaning that the associations are unique to individuals (Ward & Simner, 2003; Deroy and Spence, 2013). Although most sources agree that synesthetic associations are random, anecdotal evidence suggests that some synesthetes' synesthetic associations are linked to particular memories from childhood, like colored refrigerator magnets (Witthoft & Winnawer, 2006).

4.2.2.2 Testing for synesthesia

Because synesthesia is a perceptual phenomenon that is unique to individuals, it was once very difficult to test in participants. All widely accepted methods of testing for synesthesia have relied upon testing either internal consistency, strength of associations, or both and require synesthetes to outperform control individuals who do not claim to have synesthesia.

Traditionally, an individual who was believed to have synesthesia was tested based on the internal consistency of their associations over time. An early standard for testing synesthesia, the Test of Genuineness (TOG-R: Baron-Cohen et al., 1987), required synesthetes to report their synesthetic response to a set of stimuli and to return months later to recall the associations they had initially reported. Synesthetes were expected to outperform controls who had been asked to perform the same task using random responses and their memories upon re-test (Ward & Simner, 2003). One particular limitation of the TOG-R was that the methodology of testing varied among studies, making comparison across studies difficult. For example, participants were asked to indicate synesthetic associations with graphemes using written accounts (Simner et al., 2006), verbal descriptions (Ward et al., 2005), online color wheels (Simner et al., 2009), the Pantonec swatch color wheel (Asher et al., 2006), and customized color selection programs with over 16 million shades (Simner & Ludwig, 2012).

A modification of the Stroop test (Stroop, 1935) was often used to test for the strength of synesthetic associations, and our lab has altered this approach (Martinez, 2018). In a Stroop test, participants view the written names of colors, which are presented with matching or congruent font colors (e.g. 'green' in green font) or mismatching or incongruent font colors (e.g. 'red' in green font). Individuals are typically slower to respond to incongruent than congruent trials due to the mismatch of colors. In potential grapheme-color synesthetes, the Stroop task is tailored to fit a participant's specific color and letter associations to measure strength but not consistency of the associations. The greater the difference between congruent and incongruent response times, the greater the interference of mismatching letter-color pairs, which implies that the synesthetic associations are stronger.

A modern, computerized, and less complex diagnostic measure of synesthesia is now commonly used called the Synesthesia Battery (Eagleman et al., 2007). This assessment measures both strength and consistency of various types of synesthetic associations, but excludes those that cannot easily be tested online, such as taste and smell. For our purposes, the grapheme-color Synesthesia Battery tests include participants' selection of the exact colors they see for each letter or number in randomized order to assess the consistency of their associations. Then, participants complete a behavioral task to measure the strength of their associations, and finally participants complete a questionnaire describing their synesthetic preferences and how they experience them. This battery has been validated in multiple studies that have illustrated a bimodal distribution between synesthetes and non-synesthetes based on performance on the Synesthesia Battery and have established that this method is comparable to more traditional methods of testing synesthesia (Carmichael et al., 2015; Eagleman et al., 2007; Rothen et al., 2013).

4.2.2.2 The nature of synesthetic experience

The Synesthesia Battery has become the most commonly used method of testing for the presence of synesthesia, as well as the type of synesthesia. Synesthetes are often classified into two types as either associators or as projectors (Dixon et al., 2004). Most synesthetes are associators, which means that their secondary experiences is in the "mind's eye" and not physically perceived. A grapheme-color synesthete who is an associator would feel certainly that a specific shade of blue goes with or is associated with the letter "K", and seeing the letter "K" brings this shade of blue to mind quickly and automatically. In contrast, a grapheme-color synesthete who is a projector experiences a perceptual reality in which they actually see the associated color as a photism sitting over the grapheme as though it were a transparency placed over the lens of a projector (Palmeri et al., 2002). These photisms experienced by synesthetic projectors can affect their perceptions in other tasks. When participants are primed with colored gratings and are then presented with colorless gratings, participants will perceive the colorless gratings as having color if they are presented in the same orientation. This phenomenon, known as the McCullough effect, is enhanced for synesthetic projectors (Ramachandran & Marcus, 2017). Distinguishing a shape constructed with many of the number "5" from within a matrix of "2"s is difficult for most individuals because of the orthographic similarity of “2” and “5.” A projector synesthete can easily identify the shape because the contrasting colored photisms for “2” and “5” accentuate the shape, which indicates that synesthetic photisms can enable perceptual grouping (Ramachandran & Hubbard, 2001).

The classification of projector versus associator brings up one of the contentious points in the literature of synesthesia, which is how to characterize the actual synesthetic experiences themselves (Marks, 2011). Across multiple studies, synesthetic experiences have been described as sensory (Ramachandran & Hubbard, 2001; 2003), perceptual (Palmeri et al., 2002; Rich & Mattingley, 2002; Segal, 1997), conceptual (Simner, 2007), or related to mental imagery (Galton, 1880; Rader & Tellegen, 1987), illusions (e.g. hallucinations or after images; Lycan, 2006; Sagiv et al., 2011), and even non-experiential visual seemings that expand the linguistic meaning of ‘to see’ (Brogaard, 2013). We here take the stance that the associations present in synesthesia are perceptual, at least in part and for some individuals, although not all other accounts of the nature of synesthetic experience are mutually exclusive to the idea that the associations are based in perception. In sum, synesthestes are a unique example of individual difference in perception that may shed important light into the question of how MSI can vary so greatly across individuals while still leading to consistent patterns of association such as sound-symbolic CCs.

4.3 Materials and Methods

4.3.1 Typical individuals

See Chapter 3, Section 3.3. for details on the neuroimaging and behavioral tasks that a group of typical individuals (i.e. with no reported synesthesia; n = 24) completed. For neuroimaging, we here focus on the multisensory session (See Chapter 3, Section 3.3.4), and for behavioral results, we here focus on the individual differences in the task completed during the multisensory session (See Chapter 3, Section 3.3.4), the congruency assay, and the post-scan behavioral testing which consisted of rating shapes and pseudowords on a scale of rounded to pointed using numbers 1-7 (for additional detail, see Chapter 3, Section 3.3.5).

4.3.2 Synesthete

One grapheme-color synesthete, hereafter referred to as TR (F; 21 years), was recruited to complete the same series of neuroimaging and behavioral tasks described in Chapter 3, Section 3.3. In addition, TR completed the Synesthesia Battery (Eagleman et al., 2007) with a score of 0.52 with scores less than one indicating a score of synesthetic, and her project/associator score was indicating that she is a strong projector. Upon interviewing, TR noted that the ratings for pseudowords were particularly difficult because of the complex color percepts she experienced when hearing the pseudowords as well as the color concurrents she received from the numbers on the keyboard. She was having difficulty keeping a scale of roundedness or pointedness in mind while attempting not to match the colors of the pseudowords with the colors of the keyboard numbers. Thus, she completed a second post-scan behavioral test that was modified such that she rated pseudowords and shapes on a scale of rounded to pointed using a vertical bar. The ratings she completed with the vertical bar were then translated into numbers 1-7 depending on the height of the bar. For example if the height of 2.1 cm was marked on a 7 cm bar, then the rating was 2.

4.4 Results

4.4.1 Typical variation in perception from behavioral analyses

Although the group-level analysis indicated that most individuals were sensitive to sound-symbolic CCs (See Chapter 3, Section 3.4.1), we sought to investigate how each individual performed on the two measures of sound-symbolic crossmodal correspondences. Figure 4.1 illustrates the distribution of the normalized congruency assay taken in the scanner, which indicates that there were differences across individuals in performance and that four individuals did not show a congruency effect.

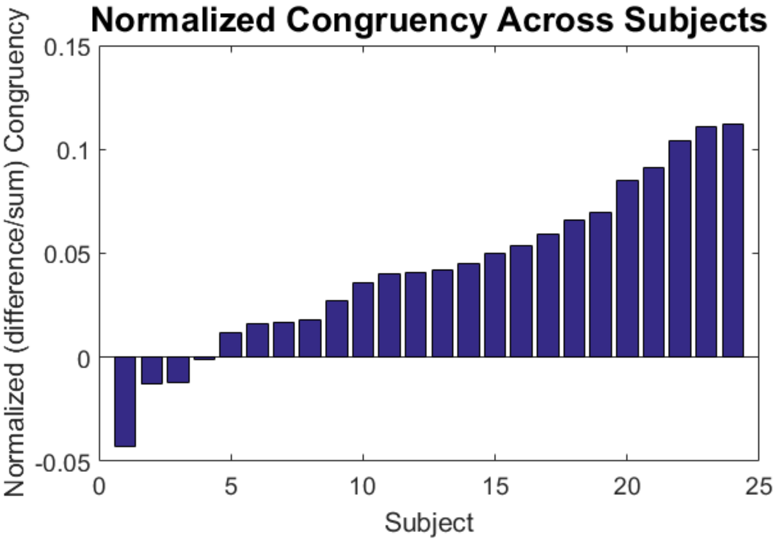


Figure 4.1. The distribution of the normalized congruency effect across participants illustrates that four participants did not demonstrate a congruency effect, although the majority were sensitive to sound-symbolic CCs.

Figure 4.2 illustrates the distribution of the second order Spearman correlation across the pseudoword and shape ratings of rounded/pointedness, which indicates that there were differences across individuals in performance and that six individuals did not show a congruency effect. We then attempted to correlate the performance on both of the behavioral assays that are meant to determine sensitivity to sound-symbolic CCs, but this correlation was not significant ($r = 0.41$; $p = 0.10$). Perhaps these two behavioral assays test different aspects of sound-symbolic CCs that are not predictive of one another. The congruency assay may be more of an implicit test, while the shape and pseudoword ratings may involve more direct and explicit judgment.

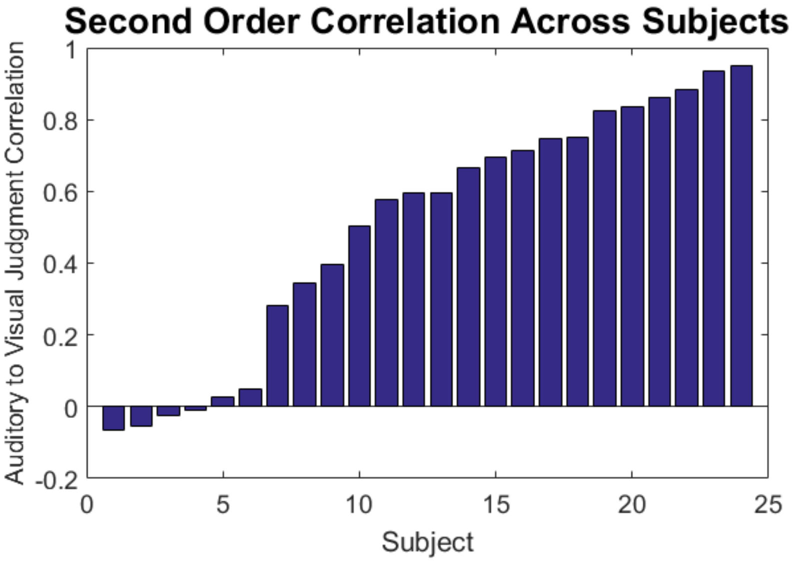


Figure 4.2. The distribution of the second order Spearman correlation for the visual perceptual rating RDMs versus the auditory perceptual rating RDMs across participants. Six participants did not demonstrate a positive correlation between the ratings across modality, although the majority was sensitive to sound-symbolic CCs.

In sum, although the group level indicated that most individuals were sensitive to sound-symbolic CCs, each individual performed differently on the two measures of sound-symbolic crossmodal correspondences, and their performance on one task did not predict their performance on another.

4.4.2 Synesthetic variation in perception from behavioral analysis

We then investigated the similarities and differences that a single grapheme-color synesthete exhibited on both the behavioral measures of sound-symbolic CCs. Although TR performed well on the multisensory congruency assay in the scanner (0.081; See Figure 4.1), we were initially surprised to find that she did not show the expected patterns in her ratings of the pseudoword rounded/pointedness (Figure 4.3, left). Upon interviewing, TR noted that the ratings for pseudowords were particularly difficult because of the complex color percepts she experienced when hearing the pseudowords as well as the color concurrents she received from the numbers on the keyboard. She was having difficulty keeping a scale of roundedness or pointedness in mind while attempting not to match the colors of the pseudowords with the colors of the keyboard numbers. We therefore modified the post-scan rating task using a vertical bar instead of numbers (Section 4.3.2) and found that her ratings matched those of typical participants when the confound of competing color percepts was removed from the task (Figure 4.3, right).

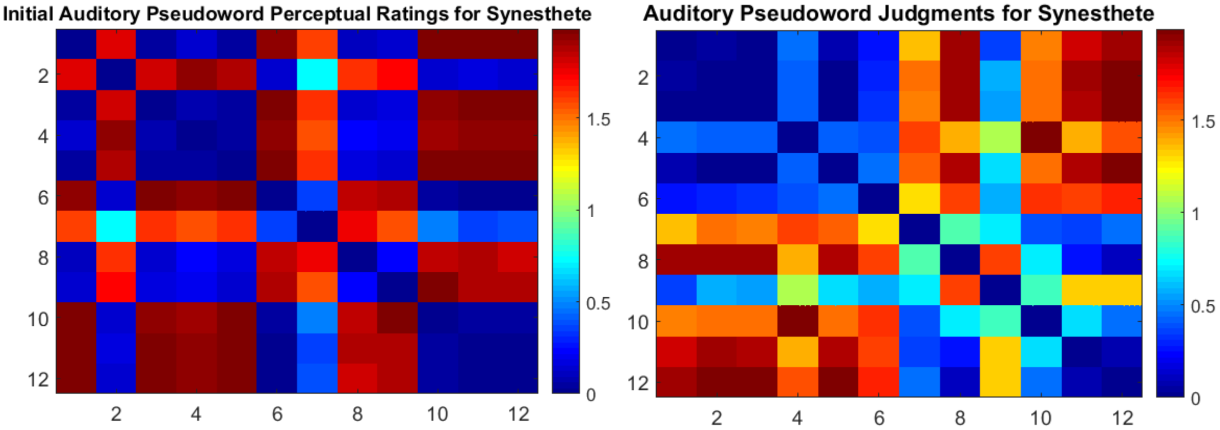


Figure 4.3. RDMs ordered according to mean perceptual judgments for auditory pseudowords in one grapheme-color synesthete. On initial testing (left), TR noted that the numeric assignment of roundedness or pointedness was distracting with the colors she saw for numbers versus pseudowords, and as a result, no correlation is evident across both sets of stimuli (second order Spearman r = 0.01, p=0.56). TR was re-testing using a modified rating procedure involving a vertical bar instead of numbers, which led to the expected pattern of ratings across shapes and pseudowords (second order Spearman r = 0.86, p<0.001).

4.4.3 Neural variation across individuals

Finally, we examine the quantitative and qualitative similarities and differences evident in one synesthete. For the quantitative differences in TR's multisensory session compared to the typical group, we first performed a univariate contrast of congruent>incongruent (Table 4.1). Like the typical individuals, the synesthete's multisensory contrast revealed multiple regions that are known to be important for auditory processing, visual processing, auditory-visual processing, and speech processing (See Chapter 3, Section 4). Listed first are the areas from the left hemisphere in the order of frontal, parietal, temporal, occipital lobe areas and then basal ganglia and cerebellar areas, and second listed are the areas from the right hemisphere in same anatomical order. Unlike the group of typical individuals, there were not more clusters on the left hemisphere than on the right hemisphere, which in typical individuals (Table 3.2) may be explained by the fact that sound-symbolic CCs involved speech sounds, and many neural correlates of language are left lateralized in the brain (Pulvermüller et al., 2006).

Region	**Cluster Size**	**x**	**y**	**z**
L Superior Frontal Gyrus	91	-18	8	56
L Middle Frontal Gyrus	1537	-46	34	18
L Superior Temporal Gyrus	5	-62	-2	-14
L Inferior Temporal Gyrus	67	-50	-50	-16
L Supramarginal Gyrus	10615	-50	-26	34
L Supplementary Motor Cortex	42	-6	-8	48
L Middle Cingulate Gyrus	425	-6	8	34
	24	-10	-24	-40
L Head of Hippocampus	32	-20	-14	-8
L Fusiform Gyrus	568	-38	-30	-20
R Superior Frontal Gyrus	73	20	8	50
R Middle Frontal Gyrus	639	44	44	6
R Frontal Operculum	163	36	18	6
R Middle Temporal Gyrus	422	56	-58	2
	33	52	-40	4
R Temporal Pole	16	52	16	8
R Supramarginal Gyrus	2466	56	-18	36
R Precentral Gyrus	40	42	2	48
R Supplementary Motor Cortex	7	8	12	50
R Middle Cingulate Gyrus	26	10	-16	42
R Lingual Gyrus	30	20	-40	-16
R Putamen	137	20	4	2
R Thalamus	43	4	-20	-2
R Lateral Cerebellum	11	46	-44	-36

Table 4.1. Multisensory fMRI session activations for synesthete: congruent > incongruent, within cortical mask, FWE-corrected $p < .05$, cluster threshold 5 voxels; x, y, z: MNI coordinates for local maxima.

After investigating the multisensory results, we next examined the unisensory visual and unisensory auditory tasks in TR with univariate analyses. As in Chapter 3, we divided the twelve stimuli in each respective modality, shapes for visual and pseudowords for auditory, such that six stimuli were categorized as a “pointed” event and the other six were categorized as a “rounded” event. The contrast of pointed > rounded in the visual session within the cortical mask (uncorrected p < .001, cluster threshold 0 voxels) revealed multiple additional areas that may be important for the perception of sound symbolic crossmodal correspondences in the visual modality, at least for one synesthete (Table 4.2). Unlike in the multisensory analysis for TR, areas such as the left postcentral gyrus, the left angular gyrus, left inferior occipital gyrus, left putamen, right superior temporal gyrus, and right postcentral gyrus are present, as well as areas that are important for language processing, such as the left lingual gyrus. As in the multisensory analysis for the synesthete, we see left precentral gyrus, left cingulate, right superior frontal gyrus, right middle temporal gyrus, right putamen, and right thalamus for both multisensory congruent>incongruent and unisensory visual pointed > rounded.

In the visual session for TR, we also examined the reverse contrast of rounded > pointed within the cortical mask (uncorrected p < .001, cluster threshold 4 voxels) that revealed several important areas (Table 4.3). Like the pointed>rounded contrast for the visual session, we see right superior frontal gyrus, right middle temporal gyrus, right superior temporal gyrus, and left lingual gyrus, but also lingual gyrus in the right hemisphere. In this contrast, we also observe left middle frontal gyrus, left inferior frontal gyrus, and right occipital fusiform gyrus for the distinction of rounded over pointed stimuli.

Region	**Cluster Size**	**x**	**y**	**z**
L Postcentral Gyrus	2	-52	-16	24
L Anterior Cingulate Gyrus	1	-6	36	20
L Middle Cingulate Gyrus	1	-2	20	24
	1	-2	-16	28
L Superior Parietal Lobule	1	-14	-74	46
L Angular Gyrus	1	-50	-54	28
	2	-32	-58	26
L Lingual Gyrus	1	-6	-86	-18
L Inferior Occipital Gyrus	2	-36	-86	-22
L Thalamus	3	0	-12	-4
	1	-20	-32	18
R Superior Frontal Gyrus (medial)	1	8	28	50
R Precentral Gyrus	1	16	-18	58
R Postcentral Gyrus	2	44	-16	30
R Middle Temporal Gyrus/Superior Temporal Gyrus	1	46	-18	-12
R Middle Temporal Gyrus	1	46	-60	14
R Putamen	1	24	2	8
R Thalamus	1	16	-26	22

Table 4.2. Unisensory visual fMRI session activations for synesthete: pointed > rounded, within cortical mask, uncorrected p < .001, cluster threshold 0 voxels; x, y, z: MNI coordinates for local maxima.

Region	**Cluster Size**	**x**	**y**	**z**
L Middle Frontal Gyrus	9	-36	40	2
L Inferior Frontal Gyrus/Fusiform	5	-48	-54	-22
L Lingual Gyrus	6	-8	-86	-22
L Lingual Gyrus/Cerebellum	6	-10	-78	-16
R Superior Frontal Gyrus (medial)	4	4	44	44
R Superior Temporal Gyrus	4	68	-30	4
R Middle Temporal Gyrus	5	60	-26	-14
R Occipital Fusiform Gyrus	4	32	-74	-12
R Lingual Gyrus	4	12	-80	-14

Table 4.3. Unisensory visual fMRI session activations for synesthete: rounded > pointed, within cortical mask, uncorrected p < .001, cluster threshold 4 voxels; x, y, z: MNI coordinates for local maxima.

The contrast of pointed > rounded in the auditory session within the cortical mask (uncorrected p < .001, cluster threshold 4 voxels) revealed multiple additional areas that may be important for the perception of sound symbolic crossmodal correspondences in the auditory modality (Table 4.4). Unlike in the multisensory analysis for TR, areas such as left angular gyrus, left middle occipital, left precuneus, and right precuneus are present. The left middle temporal gyrus is also present, whereas the same area appears in the right for the multisensory contrast. As in the multisensory analysis for the synesthete, we see left superior frontal gyrus, middle frontal gyrus, right superior frontal gyrus, and right middle frontal gyrus in the pointed > rounded contrast for the auditory session.

In the auditory session, we also examined the reverse contrast of rounded > pointed within the cortical mask uncorrected at p<.001 with threshold of 4 voxels (Table 4.5). Unlike in the multisensory analyses, areas such as the left precentral gyrus and left anterior insula are present.

As in the multisensory analyses, we see left supramarginal gyrus in the auditory rounded > pointed contrast.

Region	**Cluster Size**	**x**	**y**	**z**
L Superior Frontal Gyrus (medial)	7	-6	58	6
L Superior Frontal Gyrus	5	-20	58	8
L Middle Frontal Gyrus	5	-42	44	16
	11	-34	14	42
	11	-30	24	46
L Angular Gyrus	4	-34	-60	38
	10	-46	-60	36
	5	-28	-68	-40
L Middle Temporal Gyrus	4	-60	-54	-2
L Middle Occipital Gyrus	8	-38	-78	26
L Precuneus	9	0	-58	24
R Superior Frontal Gyrus (medial)	8	8	60	0
R Middle Frontal Gyrus	8	38	28	38
	7	48	28	30
	6	40	40	22
R Precuneus	6	2	-58	14

Table 4.4. Unisensory auditory fMRI session activations for synesthete: pointed>rounded, within cortical mask, uncorrected p < .001, cluster threshold 4 voxels; x, y, z: MNI coordinates for local maxima.

Region	**Cluster Size**	**x**	**y**	**z**
L Precentral Gyrus	7	-40	2	26
L Anterior Insula	5	-36	20	2
L Superior Parietal Lobule/Supramarginal Gyrus	6	-30	-42	36

Table 4.5. Unisensory auditory fMRI session activations for synesthete: rounded>pointed, within cortical mask, uncorrected p < .001, cluster threshold 4 voxels; x, y, z: MNI coordinates for local maxima.

Finally we examine the qualitative differences between the synesthete, one typical individual, and the entire typical group during the multisensory session with a univariate contrast of congruent > incongruent (FWE $p < 0.05$). The single participant was selected for their positive performance on both behavioral assays for sensitivity to sound-symbolic CCs. The figure panels include a sagittal (top left), coronal (top right), and transverse (bottom left) view. Within the transverse panel, right is the anterior direction, and left is the posterior direction. Within the coronal panel, left is the left hemisphere, and right is the right hemisphere. The map of the synesthete largely overlaps with that of the single individual and the group, particularly in the sagittal view of the left hemisphere. The synesthete exhibits more anterior and posterior spread than the group, which may be consistent with the idea that synesthetes generally have more connections leading to auditory-visual crossmodal processing (Brang et al., 2012).

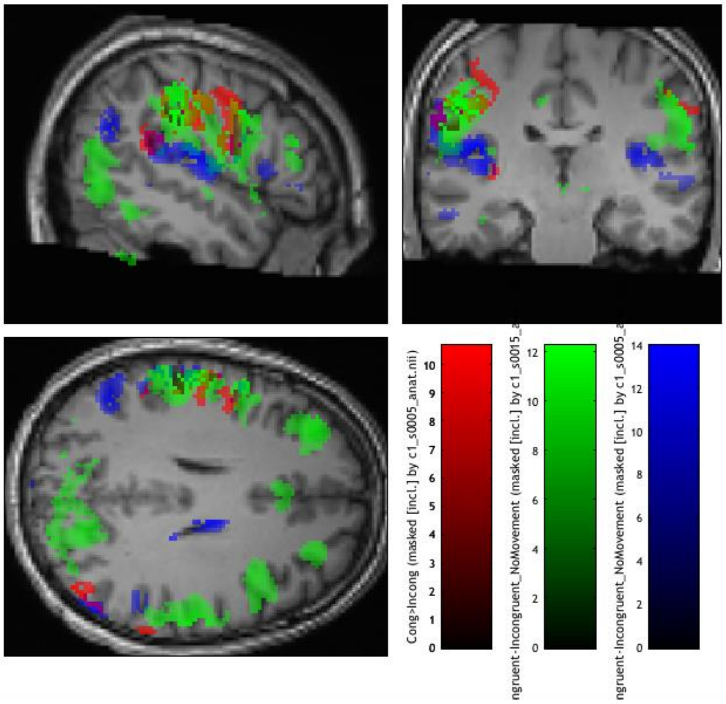


Figure 4.4. Qualitative comparison of the synesthete (green), one participant (blue), and the second level group analysis (n = 24; red) on the multisensory session with a univariate contrast of congruent > incongruent (FWE $p < 0.05$). Panels include a sagittal (top left), coronal (top right), and transverse (bottom left) view. Within the transverse panel, right is anterior, left is posterior. Within the coronal panel, left is left hemisphere, and right is right hemisphere. The map of the synesthete largely overlaps with that of the single individual and the second level group analysis, particularly in the sagittal view of the left hemisphere. The synesthete exhibits more anterior and posterior spread than the group.

Another view of the overlap across these maps supports the crossmodal hypothesis (Figure 4.5) in that from the coronal and sagittal panels, both the individual participant and the synesthete, who each performed well on both measures of sound-symbolic CCs, exhibit more bilateral activity than the group map. In addition, the synesthete exhibits more right insula activity, which is particularly visible from the transverse panel.

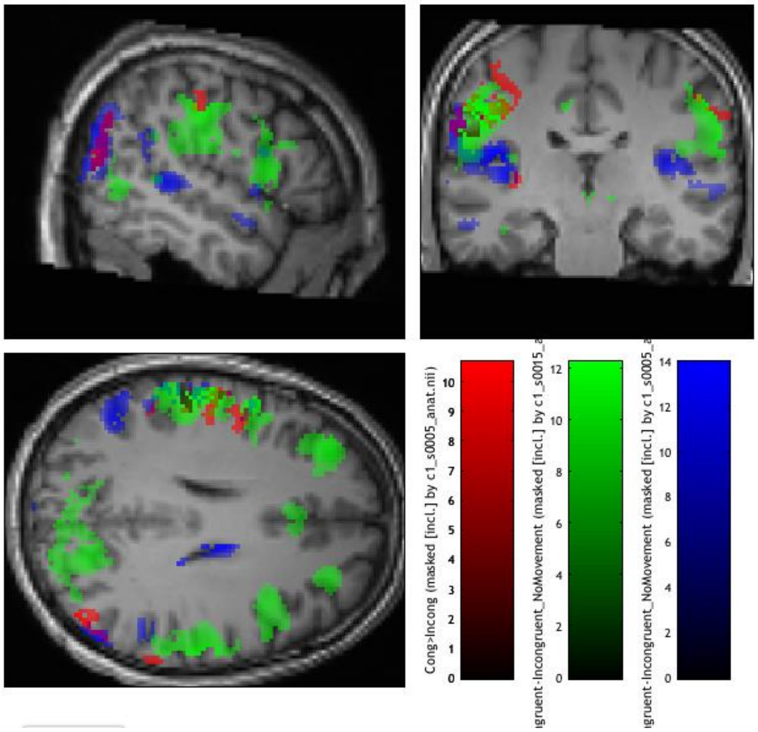


Figure 4.5. A second view of the qualitative comparison of the synesthete (green), one participant (blue), and the second level group analysis (n = 24; red) on the multisensory session with a univariate contrast of congruent > incongruent (FWE $p < 0.05$). Panels include a sagittal (top left), coronal (top right), and transverse (bottom left) view. Within the transverse panel, right is anterior, left is posterior. Within the coronal panel, left is left hemisphere, and right is right hemisphere. The synesthete exhibits more right insula activity, particularly visible from the transverse panel. From the coronal and sagittal panels, both the individual participant and the synesthete, who each performed well on both measures of sound-symbolic CCs, exhibit more bilateral activity than the group map.

A third view of the overlap among the synesthete and typical individuals reveals evidence toward the nature of this particular synesthete's color percepts (Figure 4.6). From the coronal and sagittal panels, the synesthete exhibits a clear and large swatch of occipital activity, which is likely a result of the color percepts that TR received from the stimuli. TR did describe color percepts to the pseudowords and even to the noise of the scanner, which she described as a shade of dark red during functional scans that included a light purple during structural scans. She maintains that these colors were "not distracting", but these reports combined with the fact that these color concurrents to sound yielded large activations in visual cortex suggest TR's synesthesia is based in perception. TR actually sees the colors that she describes for sounds and graphemes, which is consistent with her SB test indicating that she is a projector.

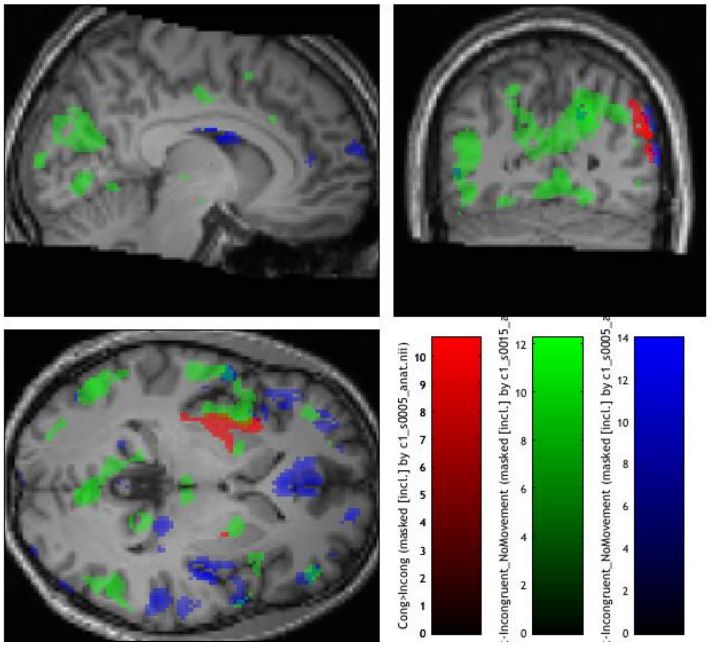


Figure 4.6. A third view of the qualitative comparison of the synesthete (green), one participant (blue), and the second level group analysis (n = 24; red) on the multisensory session with a univariate contrast of congruent > incongruent (FWE $p < 0.05$). Panels include a sagittal (top left), coronal (top right), and transverse (bottom left) view. Within the transverse panel, right is anterior, left is posterior. Within the coronal panel, left is left hemisphere, and right is right hemisphere. From the coronal and sagittal panels, the synesthete exhibits a clear and large swatch of occipital activity, which is likely a result of the color percepts that TR received from the stimuli.

In sum, the grapheme-color synesthete exhibited evidence for her concurrents of color to be based in perception (Figure 4.6) but also demonstrated that much of TR's sound-symbolic CC activations overlapped with that of the group of typical individuals and even with the map of one individual who was sensitive to sound-symbolic CCs.

4.5 Discussion

We have here demonstrated that even within the same experimental context, multiple individuals can perceive and subsequently respond quite differently to sound-symbolic CCs. Although the group level analysis indicated that most individuals were sensitive to sound-symbolic CCs (See Chapter 3, Section 3.4.1), we investigated how each individual performed on the two measures of sound-symbolic crossmodal correspondences and found that each individual performed differently on the two measures of sound-symbolic crossmodal correspondences, and their performance on one task did not predict their performance on another. We then investigated the similarities and differences that a single grapheme-color synesthete exhibited on both the behavioral measures of sound-symbolic CCs and found she was having difficulty keeping a scale of roundedness or pointedness in mind while attempting not to match the colors of the pseudowords with the colors of the keyboard numbers. We therefore modified the post-scan rating task and found that her ratings matched those of typical participants when the confound of competing color percepts was removed from the task. Finally, we examined the qualitative similarities and differences evident in one synesthete, one participant who performed well in the behavioral assays of sound-symbolic CCs, and the second level group analysis (n = 24) during the multisensory session with a univariate contrast of congruent > incongruent (FWE $p < 0.05$). We found that the grapheme-color synesthete exhibited evidence for her concurrents of color to be based in perception (Figure 4.6) but also demonstrated that the majority of TR's sound-symbolic CC activations overlapped with that of the group of typical individuals and even with the map of one individual who was sensitive to sound-symbolic CCs. These findings may support the theory that synesthetes have increased processing of auditory and visual modalities that extends to increased sensitivity to sound-symbolic CCs (Brang et al., 2012; Lacey et al., 2016).

Because we are comparing one individual to a group, we have not made direct quantitative comparisons between the activations for a given contrast between TR and those of the group. Nevertheless, there were several areas that overlapped between the synesthete and our typical group for the multisensory congruent>incongruent contrast. For example, in the unisensory visual pointed>rounded contrast for TR, the left postcentral gyrus (MNI coordinates of local maximum -52, -16, 24) was within the left precentral gyrus-postcentral ROI from the multisensory contrast of congruent>incongruent. The areas of the left angular gyrus (MNI coordinates -50, -54, 28) from the pointed>rounded in the visual session for TR were adjacent to the left precentral gyrus-postcentral gyrus ROI from the multisensory session. Because sound symbolic CCs are an association of specific acoustic properties of speech onto a reference of specific visual features, it is perhaps even more interesting that there were several areas in the unisensory auditory pointed>rounded and rounded>pointed contrast for TR that were adjacent to or near the left precentral gyrus-postcentral gyrus ROI, the left middle frontal gyrus ROI (MNI coordinates -32, 40, 22), and the left ventral angular gyrus (MNI coordinates -36, -46, 34) ROI that were present for the multisensory contrast congruent>incongruent for the typical group. These areas that are adjacent for the pointed>rounded contrast include the left middle frontal gyrus (MNI coordinates –30, 24, 46) and the left angular gyrus (MNI coordinates -46, -60, 36). The area of the left precentral gyrus (MNI coordinates -40, 2, 26) in the auditory session contrast of rounded>pointed for TR is adjacent to the precentral gyrus-postcentral gyrus cerebral cortical region from the multisensory congruent>incongruent contrast for the typical group.

This work questions the sufficiency of group descriptions for scientific studies, especially for research that involves multisensory processing, which as we have shown, is highly variable across individuals and dependent on the task or context. We here suggest that individual variation in our results adds to the ecological validity of the findings because it replicates a general principle of the field of multisensory integration. Further, by including one synesthete in our experiments, we have illustrated the similarities and differences that exist among individuals in a particularly extreme case of perceptual variation.

4.6 Conclusion

Sound-symbolic CCs may implicate auditory-visual multisensory integration, a process by which the input from multiple sensory modalities is combined into a single coherent consciousness. Although sound-symbolic crossmodal correspondences are consistent across language background, cultural background, and children pre-reading age, one of the fundamental principles of MSI is that it is context-dependent. Here, we expand on experiments presented in Chapter 3 and explore this purported paradox between a consistent phenomenon that is supported by a highly flexible neural process. We have here demonstrated that even within the same experimental context, multiple individuals can perceive and subsequently respond quite differently to sound-symbolic CCs behaviorally. These differences are evident neurally in the qualitative similarities and differences evident in one synesthete, one participant who performed well in the behavioral assays of sound-symbolic CCs, and the second level group analysis (n = 24) during a multisensory task. These findings provide preliminary evidence for the theory that synesthetes have increased processing of auditory and visual modalities that extends to increased sensitivity to sound-symbolic CCs. This research provides insights into the fundamental nature of CCs and they function within the context-dependent nature of MSI.

Chapter 5

General Conclusions and Discussion

Previous chapters have delineated the psychophysical and neurobehavioral studies through which the author has examined sound-symbolic crossmodal correspondences and what this phenomenon can reveal about sensory influences on language processing and about auditory-visual multisensory integration more generally. In this final chapter, the author summarizes and integrates the findings of the previous three chapters in the context of the project as a whole. We have investigated the physical, neural, and perceptual parameters of sound-symbolic crossmodal correspondences, leading to three main conclusions. In addition, the author considers the contribution of this project to the general knowledge of the neural and behavioral basis of sound-to-meaning mappings like sound-symbolic crossmodal correspondences and proposes future studies that will continue this work.

5.1 Summary of Findings

While others have examined visual (Chen et al., 2016) or acoustic properties (Knoeferle et al., 2017) that are important for sound-symbolic CCs, in Chapter 2, we used representational similarity analysis (See Chapter 1, Section 1.6) to examine the relationships between physical stimulus parameters and perceptual ratings that varied on dimensions of roundedness and pointedness, for a range of auditory pseudowords and visual shapes (See Chapter 2; List et al., BioRxiv; McCormick et al., 2015; McCormick et al., BioRxiv; McCormick et al., unpublished data). The systematically developed and large stimulus set (90 shapes and 537 pseudowords) of McCormick and colleagues is more extensive than those of others who have explored similar questions (See Appendix 1 & 2; McCormick et al., 2015; McCormick et al., BioRxiv;

McCormick et al., unpublished data). This advantage allowed us to explore a spectrum of rounded to pointed shapes and pseudowords. In our study, we compared the relative influence of acoustic and visual parameters on sound-shape judgments. First, McCormick et al. (unpublished data) systematically constructed visual shapes (McCormick et al., unpublished data) to assess the role of visual features in sound-shape symbolic judgments (List et al., BioRxiv), and second, McCormick et al. (2015) constructed auditory pseudowords (McCormick et al., 2015; McCormick et al., BioRxiv) to assess the role of acoustic features in the same set of judgments. (List et al., BioRxiv). Our study detailed in Chapter 2 (List et al., BioRxiv) is also the first to investigate the similarities in perceptual judgments of roundedness/pointedness for shapes compared to pseudowords and to further use these similarities to determine which visual and acoustic properties are important for sound-symbolic CCs both within and across modalities.

The findings of Chapter 2 suggest that sound-symbolic crossmodal correspondences are driven by specific acoustic and visual parameters. In addition, the parameters that predict judgments within a modality are not necessarily the same as those that predict judgments crossmodally. Specifically, simple matching coefficient, silhouette, and Jaccard distance predicted judgments of the roundedness/pointedness of shapes, while shimmer, frequency FFT, mean HNR, speech envelope, and spectral tilt predicted sound- symbolic judgments of the roundedness/pointedness of pseudowords. Crossmodally, only FFT for pseudowords and simple matching coefficient for shapes predicted judgments of roundedness/pointedness across both auditory pseudowords and visual shapes. Individuals are not only sensitive to sound-symbolic CCs in a multisensory context including both shapes and sounds, but are also able to employ these associations based on pseudowords even in the absence of the visual shapes that the sounds represent. These findings suggest that sound symbolism is not only a set of visual categorical contrasts that are instantiated by phonetic properties of words, but individuals are instead able to base their sound-symbolic judgments on a continuum of visual and auditory properties. This finding is important because it may shed light on the theoretical explanations of sound symbolism, which serves to give insight into why we have sound-symbolic CCs at all.

The question of whether individuals base their sound-symbolic judgments on a continuum (e.g. a range of sounds to symbolize the spectrum of brightness from dim to medium to bright) or categorical opposites (e.g. contrasting phonetic categories to symbolize dim versus bright) is cited as potential evidence for two theories that are meant to explain the evolution of sound symbolism. This question is known as the continuous-contrastive marking problem (Thompson & Estes, 2011) and is central to the theories on the evolutionary development of sound symbolism in language. The statistical theory states that the evolution of sound symbolism evolved because listeners consistently registered the random distribution of sounds in language. Certain sounds became statistically more likely to be present in words with specific meanings by chance, such as the phonological differences between English nouns and verbs (Farmer, Christiansen, & Monaghan, 2006). This theory most readily explains contrastive sound-to-meaning mappings in that the pairing between a sound and its meaning is initially arbitrary but over time becomes symbolic because of the statistical co-occurrence of those sounds (Thompson & Estes, 2011).

In contrast, the crossmodal theory states that the evolution of sound symbolism was spurred by listeners matching gestures and acoustic properties of speech to physical properties of stimuli in vision or other sensory modalities (Ramachandran et al., 2001). Gestures, acoustic properties of speech sounds, and physical properties in vision or other sensory modalities can all be graded and finely tuned to portray a continuous spectrum of meaning. For example, Thompson and Estes (2011) found that the number of large-sounding phonemes in an object label was associated with object size and dimensions. Knoeferle et al. (2017) extended these findings in support of the crossmodal theory by expanding from sound-size to sound-shape symbolism and examined acoustic properties of vowels. They found that sound-symbolic crossmodal judgments varied linearly as a function of the second and third formants and that different acoustic properties were associated with the two different meaning dimensions. Future studies will need to explore the acoustic and visual parameters discussed here, among others, in the context of multiple dimensions of sound-symbolic meaning to determine which parameters, if any, may be important for sound symbolism in general versus for a specific dimension of meaning.

We here provide a key link between continuous acoustic properties and continuous visual properties and show that the perception of roundedness and pointedness along a spectrum in the visual and auditory domains is correlated with similar variations along a spectrum in the simple matching coefficient for visual shapes and the FFT for auditory pseudowords. This finding may support the crossmodal theory that sound symbolism resulted from links across a continuous spectrum for certain auditory and visual properties. We can only speculate on the evolutionary development of sound symbolic crossmodal correspondences, but the findings of Chapter 2 provide evidence for the crossmodal theory of sound symbolism. Further, these findings imply that the relationship between words and their meanings can be non-arbitrary even at the level of auditory and visual perception and highlights the importance of sound symbolism in natural language.

Although sound-symbolic crossmodal correspondences have been studied for almost a century, few have systematically examined the cortical substrates of both the auditory and visual dimensions that elicit sound-symbolic CCs (McCormick et al., BioRxiv; Peiffer-Smadja & Cohen, 2019). In Chapter 3, to investigate the neural underpinnings of sound-symbolic CCs, we conducted a study designed to systematically determine how auditory, visual, and multisensory perception contribute to the perception of sound-symbolic crossmodal correspondences using univariate and multivariate analysis methods (See Chapter 1, Section 1.6; See Chapter 3). Specifically, we examined whether differential activation in a key set of cerebral cortical areas reliably predicted sensitivity to sound-symbolic crossmodal correspondences. In our study (See Chapter 3), we examine with both univariate and multivariate methods the influence of multiple cortical areas during auditory and visual perception of the component stimuli that elicit sound-symbolic crossmodal correspondences. First, we confirmed participants' sensitivity to sound-symbolic crossmodal correspondences with two independent behavioral tasks, one inside and one outside the scanner. Then, we identified cerebral cortical regions that are correlated with the multisensory and unisensory perception of sound-symbolic crossmodal correspondences. Finally, we performed multivariate analyses (See Chapter 1, Section 1.6) in these regions on neural data collected during the perception of visual shapes and separately during the perception of auditory pseudowords to assess the roles of visual perception and auditory perception, respectively, in sound-symbolic crossmodal correspondences.

To summarize the findings of Chapter 3, a univariate contrast of congruent > incongruent while participants perceived sound-symbolic CCs in the scanner revealed multiple areas, primarily in the left hemisphere but also in the right, that may be important for the processing of sound-symbolic CCs, for speech or language processing, for sensory processing, and even auditory-visual multisensory integration more generally. Our univariate contrasts of pointed>rounded and rounded>pointed in the unisensory sessions revealed adjacency between the left precentral gyrus-postcentral gyrus area shown by the contrast of congruent>incongruent from the multisensory session. This finding indicated that small cortical areas distinguishing roundedness/pointedness in either the auditory or visual domain is spatially contiguous with areas distinguishing a multisensory match or mismatch. Further, in correlational univariate analyses, we found activity in the bilateral supramarginal gyrus (SMG) in the visual session and left superior occipital gyrus (SOG) in the auditory session that was correlated with perceptual ratings of visual shapes or auditory pseudowords, respectively. This finding was unusual, given that a function of bilateral SMG is auditory sensation, and the primary function of the left SOG is visual, which indicates that even in the absence of one sensory modality, the perception of sound symbolic crossmodal correspondences may involve both vision and audition. Our multivariate analysis using RSA to compare perceptual ratings to unisensory perception of both auditory pseudoword and visual shapes did not reveal regions that passed tests of significance, but this finding does not preclude the possibility of multiple brain regions having a modulatory role in unisensory perception of pseudowords or shapes. These modulatory functions in the respective unisensory domains may then contribute to their multisensory role in the perception of language, particularly in the perception of sound-symbolic CCs. As evidenced by the overlap between multisensory and unisensory sessions in our univariate analyses, some brain regions, such as the left precentral gyrus-postcentral gyrus, may be important for multisensory integration the auditory-visual processing of sound-symbolic CCs. This large area from the multisensory congruent>incongruent contrast has significant functions in sensory and motor systems, so one may question why an area involved in sensory and motor systems may be involved in the processing of sound symbolic CCs. We have previously noted that in work examining the sound-symbolic associations of pseudowords and size, Sapir (1929) proposed that either certain vowels have greater acoustic volume than others in the auditory domain or the large spatial relationships between part of the articulatory apparatus during sound production (i.e., tongue position and resonance cavity) is symbolic of a larger reference. For example, a listener may perceive the pseudoword “mal” as larger than the pseudoword “mil” because the configuration of the mouth and throat is larger when producing the “a” versus the “i” sounds. While the example of sound and size articulation as an explanation for sound symbolism may be more intuitive than an equivalent explanation for sound and shape, the phonetic category of place of articulation has been shown to be correlated with judgments of roundedness/pointedness (McCormick et al., 2015). Further, mappings between articulatory configurations of the mouth and the meaning of sounds have been found for domains as abstract as that of affective state (e.g. meanings of “approach” and “avoidance” mapping to swallowing and spitting; Topolinski et al., 2014), so it is clear that the articulatory explanation for sound symbolism is not limited to sound-size mappings. For our experiments, the part of motor cortex that has been shown to map articulatory features of speech sounds (Pulvermüller et al., 2006) overlaps with the left precentral-postcentral gyrus area from the congruent>incongruent contrast. This cluster also extended into areas that have been implicated in subvocal articulation (Price et al., 2009). These findings may then support the articulatory explanation for sound symbolic CCs. These findings also suggest that the perception of sound-symbolic CCs is complex and multimodal but that canonically multisensory regions (e.g. the posterior superior temporal gyrus) may not explicitly regulate these associations in a unisensory context alone. This research provides insights into the fundamental nature of CCs and how they might evoke specific interpretations of physical meaning in natural language.

Further, RSA is a multivariate analyses tool that has many diverse uses in the field of neuroscience, particularly in examining effects that may be small and spatially or directionally variable across participants (Devereux et al., 2013). We have used a combination of multivariate and univariate approaches to objectively define regions that exhibit MSI during the processing of sound-symbolic CCs. Using a combination of these analytical approaches allows us to determine spatial patterns of activity that are either shared (univariate methods) or potentially unique (multivariate and correlational univariate methods) across our participants. The challenge of quantifying multisensory and crossmodal responses in human fMRI is that these brain imaging methods involve determining the appropriate statistical criterion for identifying multisensory interactions throughout the brain (Laurienti et al., 2005; Noppeney, 2012). Translating the established principles of MSI, which were defined on the basis of single-unit recordings in cats (Stein & Meredith, 1993), to behavioral and population-level neural responses in humans is no small task. Researchers particularly debate over how to differentiate among and interpret supra-additive and sub-additive nonlinear responses, which are multisensory responses that are greater or less than the sum of the unisensory responses, respectively (Murray et al., 2018; Noppeney, 2012; Stevenson et al., 2014). For fMRI, there is still no consensus on the best statistical criterion to identify multisensory brain activity, but certain individuals recommend different methods from using a non-linearity criterion (Calvert, 2001), to comparing multisensory responses to the mean or maxima of unisensory responses (Beauchamp, 2005), to using the criterion of inverse effectiveness (James & Stevenson, 2011). To circumvent the criterion debate entirely in determining a truly multisensory response with fMRI, we used a multisensory session with congruent and incongruent conditions. We further attempted to combine univariate and multivariate analytical approaches. We here do not rule out the possibility that there are other multivariate approaches, such as searchlight classifiers, that may yield insight into sound symbolic CCs. By using both sets of approaches, we attempted to infer a more complex relationship among unisensory auditory, unisensory visual, and multisensory aspects of sound-symbolic CCs than has previously been investigated.

Building upon experiments from Chapter 3, in Chapter 4 we investigated the behavioral and neural evidence of variation in the perception of sound-symbolic CCs. Sound-symbolic CCs rely on auditory-visual multisensory integration, a process by which the input from multiple sensory modalities is combined into a single coherent percept. Although sound-symbolic crossmodal correspondences are consistent across language background, cultural background, and in children of pre-reading age, one of the fundamental principles of MSI is that it is context-dependent. In Chapter 4, we expanded on experiments presented in Chapter 3 and used the findings to explore the paradox between the consistent phenomenon of sound-symbolic CCs on the one hand and the highly flexible neural process of MSI on the other. We first explored the variation present in a group of typical individuals (n = 24) who have all completed the neural and behavioral paradigms detailed in Chapter 3, Section 3.3. We then assessed one individual with grapheme-color synesthesia in order to explore the degree to which sound symbolic CCs vary across individuals and to speculate on the validity of the idea that sound-symbolic CCs arise from grounded cognition (Barsalou et al., 2008) or pervasive synesthetic associations (Ramachandran & Hubbard, 2001).

The findings of Chapter 4 demonstrated that even within the same experimental context, multiple individuals can perceive and subsequently respond quite differently to sound-symbolic CCs behaviorally. These differences are evident neurally in the qualitative similarities and differences evident in one synesthete, one participant who performed well in the behavioral assays of sound-symbolic CCs, and the second level group analysis (n = 24) during a multisensory task. These findings may support the theory that synesthetes have increased processing of auditory and visual modalities that extends to increased sensitivity to sound-symbolic CCs (Lacey et al., 2016). Interestingly, despite the additional perceptual experiences that this synesthete had, there were striking similarities between the synesthete and typical individuals, both in cerebral cortical BOLD signal and in behavioral results. For the synesthetic participant, the phenomena of sound-symbolic CCs was robust in the presence of simultaneous perception of sound- or shape-dependent colors that the synesthete reported, and this finding may suggest that the context in which an individual experiences a sound-symbolic CC is not the only driving factor for the perception of a strong association. In other words, the potential distraction of additional color percepts did not prevent the synesthete from perceiving an association between our chosen pseudowords and shapes and may make some sound symbolic CCs stronger. However, in certain contexts, such as the behavioral perceptual ratings, the synesthetically perceived colors did have an effect on behavioral measures before we made accommodations in the task for the synesthete. This research provides insights into the fundamental nature of CCs and how they function within the context-dependent nature of MSI.

In order to delineate how crossmodal correspondences fit within the framework of MSI, Spence and Deroy (2013) posit that crossmodal correspondences may be considered automatic upon meeting some of the four distinct criteria of non-consciousness, load-insensitivity, goal-independence, and speed. The non-conscious criterion may be defined as being pre-attentive or being able to occur without conscious control or volition once attention is established. The load-insensitivity criterion stipulates that a process is automatic if it is not hindered when the cognitive load is increased, for example during a dual-task. The goal-independence criterion eliminates strategic processes and tasks in which the person engages with a particular intention or goal over which the person exhibits control over whether the phenomenon is observed. Finally, the speed criteria states that a task is more likely to be automatic if it is a process that is demonstrated to affect the very earliest stages of information processing (i.e. is done very quickly, suggesting it is done without conscious thought). Spence and Deroy (2013) do not note definitely whether all four of these criteria or which ones must be reached in order to achieve the classification of "automatic" but rather indicate that these criteria are meant to be used to assess the manner by which any given experiment supports how these criteria relate to one another, since most experiments only address one or some of the criteria.

While our experiments did not explicitly investigate the automaticity of sound symbolic crossmodal correspondences, we can speculate over what each of the preceding chapters may indicate about automaticity. For example, Chapter 2 indicates that there are specific acoustic and visual parameters that individuals use to judge the roundedness/pointedness of pseudowords and shapes. These parameters are unlikely to be consciously parsed via a goal-directed action, and the detection of different parameters to lead to a decision of “pointed” or “rounded” may be non-conscious. This idea is particularly intriguing in the case of the pseudowords as opposed to the shapes because participants judged both sets of stimuli on a visual scale (i.e. rounded to pointed), and the fact that the participants as a group were able to create consistent criteria on which to judge pseudowords on a visual scale is unusual and perhaps suggesting a degree of non-conscious behavior.

Considering Spence and Deroy’s (2013) criteria in Chapter 3, we may consider the function of the neural regions that appeared consistently across modality and across task and whether the robustness of those regions would indicate some measure of automaticity of sound symbolic crossmodal correspondences. For instance, Figure 3.9 illustrates the large left precentral gyrus-postcentral gyrus cerebral cortical region from the multisensory contrast congruent>incongruent and how consistently this cerebral cortical region responds to various univariate contrasts during the perception of sound symbolic CCs. Not only are there positive contrast estimates for the pointed>rounded contrast for this cerebral cortical region in both visual and auditory unisensory tasks, but multiple clusters that appeared for the pointed>rounded contrast were near or within this precentral gyrus-postcentral gyrus cerebral cortical region (Figure 3.6). For example, in the unisensory visual pointed>rounded contrast, the left precentral gyrus (MNI coordinates -40, -12, 30), the left ventral precentral sulcus (MNI coordinates -38, -30, 36), two clusters in the left postcentral gyrus (MNI coordinates -46, -16, 52 and -46, -14, 28), and two clusters in the supramarginal gyrus (-50, -30, 42 and -42, -30, 36) were adjacent to the left precentral gyrus-postcentral gyrus cerebral cortical region from the multisensory contrast of congruent>incongruent. The areas of the left postcentral gyrus (MNI coordinates -40, -18, 46) and the left postcentral sulcus (MNI coordinates -46, -32, 46) from the pointed>rounded in the visual session were within the left precentral gyrus-postcentral gyrus cerebral cortical region from the multisensory session. Because sound symbolic CCs are thought to be an association of specific acoustic properties of speech onto a reference of specific visual features (See Chapter 2), it is perhaps even more interesting that there were several areas in the unisensory auditory pointed>rounded contrast that overlapped or were adjacent to the precentral gyrus-postcentral gyrus cerebral cortical region. These areas that are adjacent include two clusters in the left precentral gyrus (MNI coordinates -50, 6, 22 and -56, 2, 38). The areas of the left frontal operculum (MNI coordinates -32, 6, 4), the left inferior precentral sulcus (MNI coordinates -26, -12, 50) were within the precentral gyrus-postcentral gyrus cerebral cortical region from the multisensory congruent>incongruent contrast. In considering automaticity, the involvement of the left middle frontal gyrus (MFG) in Peiffer-Smadja & Cohen’s (2019), McCormick et al.’s (BioRxiv), and our multisensory study (Chapter 3) may speak to the role of attention in perceiving sound symbolic crossmodal correspondences. The consistent involvement of frontal areas like left MFG across studies may suggest that the criterion of the effect being non-conscious in order to be automatic is not fully met in the case of sound symbolic crossmodal correspondences (Spence & Deroy, 2013).

However, the fact that this large area in left precentral gyrus-postcentral gyrus was closely flanked by or contained several clusters that distinguished pointed over rounded stimuli in two separate modalities suggests that the nature of sound symbolic CCs may be in some measure automatic. The idea of automaticity may further be supported by the fact that the unisensory tasks, rare repeats tasks, were entirely different from the multisensory task, a congruency matching task. In other words, evidence of sound symbolic crossmodal correspondences in an individual was goal-independent. We did not explicitly examine the criterion of load-insensitivity, but different tasks (e.g. perceptual rating task, rare repeats task, congruency task) require different levels of attentional load. The fact that we were able to observe evidence of individuals' sensitivity to sound symbolic crossmodal correspondences in all three of these tasks, either behaviorally or in the cortical areas that were correlated with the task, speaks to the robustness of the phenomenon.

Chapter 4 speaks to Spence & Deroy's (2013) criteria in that we are able to see significant individual variation in the degree to which individuals are sensitive to sound symbolic crossmodal correspondences. When considering automaticity, this individual variation is perhaps most interesting from the standpoint of the criteria of speed. Our congruency task, a variant of the oft-used implicit association task (Parise & Spence, 2012; Lacey et al., 2016), indicated that participants, at both the level of the group average and the level of a majority of individuals, were faster to respond to respond to matching or congruent pairs of audio-visual stimuli than they were for mismatching or incongruent pairs of audio-visual stimuli. This finding may be seen to provide evidence for the speed criterion (Parise & Spence, 2012) in characterizing the automaticity of sound symbolic crossmodal correspondences. However, Spence and Deroy (2013) state that a difference of greater than 15 ms for audition or 80 ms for vision is needed to provide evidence for automaticity. For our experiments, the difference in response time for congruent versus incongruent trials was an average of 98 ms, and it should be noted that Spence and Deroy (2013) posit that a larger difference in response times is a weak argument to support the speed criteria of automaticity. By those standards, our experiments do not pass the speed criteria. In addition, we found that the additional perceptual experience of having grapheme-color synesthesia could alter the manner in which one individual was able to complete our tasks, particularly the perceptual judgment task.

A critical question in the field of sound symbolic crossmodal correspondences, and in fact crossmodal correspondences in general, is why these particular associations exist. The criteria set Spence and Deroy (2013) is by no means the only manner by which to classify crossmodal correspondences, and others have suggested that (Parise & Spence, 2009; Spence & Deroy, 2016) a significant challenge is to determine whether different types of crossmodal correspondences function similarly, and if not, whether these crossmodal correspondences can even by classified as one of the same type of phenomenon. The reason we here specifically point out automaticity among the set of studies in this dissertation is to highlight the possibility that this body of work may extend our understanding of MSI and the evolution of language.

If sound symbolic crossmodal correspondences are indeed automatic, perhaps the effects of this phenomenon are at least in part perceptual and not just decisional (Gallace & Spence, 2006). The Bayesian theories of MSI (Ernst & Banks, 2002) have been used to suggest that audiovisual crossmodal correspondences function as coupling priors, or pairs of stimuli that are integrated automatically (Spence, 2011). “Automatically” in this example is more loosely defined as being unintentional and happening outside of the control of the conscious observer. Our experiments follow the suggestions of a recent review of the field of language by Hasson & Willems (2018), which suggests that while the language network has been an important framework to begin to understand the neurobiology of language, that we should begin to take greater steps toward integrating the functions of language with other functions, such as cognition and perception. Also in accordance with Hasson and Willems (2018), we further acknowledge that the findings here are context-dependent, and our work contains the fallacy of most laboratory studies investigating language in that we have divorced our phenomena of interest from the natural language and communicative conversations from which is was derived. However, this fallacy is not limited to studies of language. Despite the limitations imposed by neuroscientific and psychophysical studies, we have presented evidence that the language-based phenomenon of sound symbolic crossmodal correspondences may be in part automatic or based in perception, and in part decisional or context-dependent. This evidence suggests the complex evolution of sound symbolic crossmodal correspondences should not be separated from the evolution of auditory-visual perception. Taken together, this dissertation suggests that if language is shaped by the brain and the neural firing patterns that drive it (Christiansen & Chater, 2008), then these non-arbitrary mappings imply that this aspect of language emerged through a complex combination of culture and learning biases that has become implicit in the structural framework of language.

5.2 Future Directions

The immediate plans to continue this work include expanding the sample size of synesthetes in Chapter 4 to be able to make inferences about grapheme-color synesthetes as a group compared to our group of typical individuals. We here tested only monolingual native English speakers, and future studies should investigate listeners' sensitivity to sound-symbolic crossmodal correspondences in native speakers of additional languages and ages to expand the applicability of the findings and provide evidence for which aspects of sound-symbolic crossmodal correspondences are a result of convention within a certain language. For investigating the physical parameters of crossmodal correspondences, experiments that intentionally vary acoustic properties such as amplitude, shimmer, mean HNR, and jitter within the same extensive stimulus set may shed more light on the relative influence of these acoustic properties on crossmodal correspondences. In addition, although evidence supports the idea that certain words are phonetically associated with semantic meaning (Köhler, 1929; Maurer et al., 2006; Nygaard et al., 2009), the specific characteristics of word structure that may correspond to multiple physical dimensions, other than rounded and pointed, are not yet known. Expanding the participant judgments from rounded and pointed to other binary dimensions for these pseudowords would further explore the consistency of sound-symbolic associations across visual dimensions, such as big and small (Sapir, 1929), or even across other modalities, such as the spatial perception of near and far.

From the perspective of multisensory integration development across a lifetime (Murray et al., 2016), studies that investigate the effects of age on the physical, neural, and perceptual parameters of sound-symbolic or other crossmodal correspondences are of particular interest. As an alternative, future work could extend the current studies by exploring additional crossmodal correspondences, such as mappings between sound and size (Jonas, Spiller, & Hibbard, 2017), taste (Wang et al., 2016) or emotion (Aryani et al., 2018). In addition, we earlier summarized evidence supporting the possibility that sound-symbolic crossmodal correspondences are automatic in some measure. However, Getz and Kubovy (2018) in their study of CCs between pitch and multiple features in the visual domain with multiple tasks concluded that even CCs that appear to be processed automatically include top-down components based in attention and task-relevance. Future studies may examine the precentral gyrus-postcentral gyrus in tasks involving natural speech or including words and phonemes of varying sound-symbolic nature to investigate how automatically speech may confer meaning, or at least visual shape, in more naturalistic paradigms. The studies conducted here could also be extended to individuals of varying language background because evidence has suggested some consistent patterns in sound symbolism across individuals with different language experience (Fort & Peperkamp, 2014; Blasi et al., 2016). Studies such as these would further illuminate the potential evolutionary role of sound-symbolic CCs.

Another important issue to address in future studies is what are the best predictors and measures of individual differences in sensitivity to sound-symbolic crossmodal correspondences and further, whether these differences can best be explained by differences in such measures as preference for object, spatial, or verbal imagery, individual differences in general skills requiring multisensory integration, or even variability in participants' phonological processing or general language skills. Two implicit questions in this body of work are why these particular sound-to-meaning mappings exist in the first place and why, if these associations were important for the evolution of language, they still exist in natural language, especially in the presence of the advantages of arbitrariness. These consistent sound-symbolic associations may reflect an ease of formation of sound-symbolic crossmodal correspondences for first or second language learners (Maurer et al., 2006; Nygaard et al., 2009; Revill et al., 2017) or perhaps these sound–to-meaning mappings improve semantic categorization or retrieval to produce faster, more efficient skilled language users (Kovic et al., 2010).

5.3 Applications of This Research

The current findings may add to the theories of the evolution of language, but research into sound symbolism also has practical applications that can serve societal purpose, even beyond the important purpose of adding to scientific knowledge. This work could be extended to apply to medical or therapeutic treatments, social media experience, advertisement, and literature. For example, the work from Chapter 2 investigating the physical parameters of sound-symbolic crossmodal correspondences can be extended to develop algorithms that detect sound-symbolic words within large set of recorded interviews. Algorithms that can detect sound-symbolic words in large corpora could be extended to sound symbolic words that convey emotion (e.g. “piss” versus “pee”; Aryani et al., 2018) and used to detect the moods of social media users or individuals under criminal investigation.

This work may also be applied to conjecture over the rewards that readers gain from literature, particularly poetry. This dissertation began with a quote from Lewis Carroll, and his poem *The Jabberwocky* (1871; 2016), and Mr. Carroll is among a large company of writers who employ sound symbolism in their work.

William Wordsworth (1910) in his poem *On the Power of Sound* references the connections between vision and audition. Throughout his poem, he uses sound-symbolic phrasing to delve into the abstract idea that sound and vision can equally impart to the hearer or seer the phases of natural life. Using onomatopoetic phrases for various sounds, Wordsworth creates imagery without the use of extensive visual cues. Wordsworth begins the poem referring to a sprit that acts as an “Organ of vision” that “Informs the cell of Hearing, dark and blind” (Stanza I).

Wordsworth implies that there exist in sound certain aspects that can function in a similar manner as vision, evoking “Images of voice” (Stanza I). Furthermore, the poet employs sound symbolism to concretize with words the connections that are evident between sight and sound, noting that “The headlong streams and fountains/ Serve Thee, invisible Spirit, with untired powers;/… They *lull* perchance ten thousand thousand flowers/ 'That' *roar*, the prowling lion's 'Here I am',/ …That *bleat*, how tender! of the dam/ Calling a straggler to her side./ *Shout, cuckoo!*--let the vernal soul/ Go with thee to the frozen zone;/ *Toll* from thy loftiest perch, lone bell-bird, *toll!*/” (Wordsworth, Stanza II, emphasis added). Wordsworth evokes the imagery of a mountain using the sounds of the mountain and various onomatopoetic words, and here the “invisible Spirit” again refers to the properties of sound that allow the reader, or perhaps the listener, to see.

While crossmodal correspondences indicate coherence between one sensory domain and another, metaphor uses language to link seemingly unrelated experiences. Lakoff and Johnson (1980) posit that metaphor is not only a linguistic or poetic device subject only to explicit and conscious intellectual manipulation, but also employs crossmodal associations. Metaphor, a device used often in literature, “involves all the natural dimensions of our experience, including aspects of our sense experiences: color, shape, texture, sound, etc. These dimensions structure not only mundane experience but aesthetic experience as well,” (Lakoff and Johnson, p. 235). The poet Malea Powell uses a fresh but apt metaphor by using the physical attributes of a kiwi fruit to describe the less concrete concept of perplexity by elaborating that "sometimes *perplexity* has exactly the *color* and *taste* of kiwi eaten with the skin on, a *fuzziness* that *hangs* on your teeth until you spit it out and pretend it's a perfectly normal thing to do" (Powell, 2012, emphasis added). Here Powell describes a concept that has no direct physical embodiment and invokes the senses of color, taste, and somatosensory cues (i.e. texture and weight) to evoke the idea of perplexity in the reader. Perhaps the connection between physical attributes and an abstract concept lies within the theory of grounded cognition and the idea that conceptual bodies within the mind are supported by sensory processes (Barsalou, 2008; Lacey et al., 2012; Pezzulo et al., 2013; Lacey et al., 2017). Interestingly, Lakoff and Johnson (1980) described the metaphoric inducer (i.e. the kiwi fruit here) and the concurrent (i.e. perplexity in this example) in a similar context to the synesthetic inducer and concurrent described in Chapter 4.

Based on the current and previous studies (Nielsen & Rendall, 2011; Fort & Peperkamp, 2014), these crossmodal connections may be reflected within the structure of languages in the form of cross-linguistic sound symbolism and be represented cortically through cross-sensory integration. Sound-symbolic crossmodal correspondences can be used to understand and create more evocative poetry or prose using the sounds of words rather than solely relying on their meanings to encourage mental simulation or create new metaphors, and in fact, non-arbitrary connections between sound and meaning may be the result of this mental simulation (Barsalou, 2008; Simmons, Hamann, Harenski, Hu, & Barsalou, 2008).

This dissertation, particularly Chapter 2, supports the possibility that non-arbitrary and systematic associations between words and the visual properties of their referents inform communicative strategies in natural language (Gasser, 2004; List et al. BioRxiv). If these non-arbitrary associations are due to connections between the various perceptual systems in the brain as indicated by Chapter 3, sound symbolism may allow for greater sensory simulation as well as a depth of abstraction in expression and communication not possible under the arbitrariness assumption (Barsalou, 2008; Gasser, 2004; See Chapter 3). In addition, these crossmodal associations may confer an advantage to communication that pure arbitrary word-referent assignment does not in daily use. Everyday examples of sound-symbolic advantage can be seen through the advertising industry, and studies have been conducted indicating that marketers can enhance certain perceptions of consumer products by ensuring that the sound symbolism of the brand name and sound or shape symbolism within the labeling or packaging establishes the desired product-related sensory associations, such as taste, texture, color, or smell of a chocolate bar (Shrum, Lowrey, Luna, Lerman, & Liu, 2012; Spence, 2012; Spence & Deroy, 2013; Spence et al., 2015; Wan et al., 2014; Wang et al., 2016).

Sound symbolism may operate on an unconscious level, allowing consumers to make judgments and associations with products or brands based on these implicit connections and without overtly evaluating the information presented (Spence, 2012; Wang et al., 2016). Shrum et al. (2012) found that these sound-symbolic properties hold across languages for participants who were bilingual in French, Spanish, or Chinese and in English. Participants preferred words in which there was congruency between the product qualities and the sound symbolism of the label. These results held when participants completed the study in their first or second language and regardless of second-language proficiency or whether the Chinese language representations were in logographic or alphabetic form. Understanding how particular sound segments are associated with particular physical dimensions (Chapter 2) and how these associations can vary across different individuals or consumer groups (Chapter 4) can have an influence on how advertisers plan campaigns for new products, encouraging buyers through their implicit associations with sound.

These findings are not only significant in the marketing world for brand development, but also provide evidence for the idea that certain sound-symbolic associations may arise from a bias of crossmodal mapping within the brain rather than solely developing through cultural exposure to a given language (Kantartzis, Imai, & Kita, 2011; Shrum et al., 2012). In fact, the implications for multisensory integration of this work (Chapter 3) and how these processes vary across individuals (Chapter 4) can spur developments in multisensory learning and the long-term effects of multisensory interactions (Powers et al., 2009; Zangenehpour & Zatorre et al., 2010) and by extension, clinical rehabilitative work (Johansson, 2012). For example, dyslexic children have been found to have a crossmodal deficit in matching visual letters to their sounds (Žarić et al., 2015), and Chen et al. (2016b) found that dyslexic individuals' crossmodal deficits may extend to other crossmodal correspondences as well, such as the association between visual size and auditory pitch. Findings such as these open the door for improved diagnostics and therapies for individuals with dyslexia (Drijvers et al., 2015). Sound-symbolic crossmodal associations have wide-reaching implications, from literature to advertising to social media to medicine, and this dissertation adds to the collaborative innovations from each of these fields.

5.4 Conclusions

The data reported in this dissertation contribute to the understanding of auditory-visual multisensory integration and the manner in which audition and vision support language via the example of sound-symbolic crossmodal correspondences. We have investigated the physical, neural, and perceptual parameters of sound-symbolic crossmodal correspondences, leading to three main conclusions. First, there are physical parameters in visual and auditory stimuli that elicit sound-symbolic crossmodal correspondences, which imply a perceptual basis to these associations. Second, the neural correlates that lead to the perception of sound-symbolic crossmodal correspondences suggest that the linguistic advantages of sound symbolism are deeply ingrained in our sensory systems. Third, physical parameters and neural correlates of sound-symbolic crossmodal correspondences are flexibly mediated by individual variation in perceptual capabilities. As a part of an ongoing multidisciplinary effort, this dissertation broadens the possibilities for understanding the neural and developmental evolution of language and can be applied to fields as niche as advertising and as broad as multisensory learning.

References

Ademollo, F. (2011). *The* Cratylus *of Plato: A Commentary*. Cambridge University Press: Cambridge, UK.

Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, *38*(1/4), 298. doi:10.12697/sss.2010.38.1-4.11.

Aiken, S. J., & Picton, T. W. (2008). Human cortical responses to the speech envelope. *Ear and Hearing*, *29*(2), 139-157. doi:10.1097/aud.0b013e31816453dc.

Akita, K. & Tsujimura, N. (2016). Mimetics. In T. Kageyama & H. Kishimoto (Eds.) *Handbook of Japanese Lexicon & Word Formation*, pp133-160. Walter de Gruyter Inc., Boston, USA.

Albers, A. M., Meindertsma, T., Toni, I., & De Lange, F. P. (2018). Decoupling of BOLD amplitude and pattern classification of orientation-selective activity in human visual cortex. *NeuroImage*, *180*, 31-40. doi:10.1016/j.neuroimage.2017.09.046.

Anderson, A. J., Bruni, E., Lopopolo, A., Poesio, M., & Baroni, M. (2015). Reading visually embodied meaning from the brain: Visually grounded computational models decode visual-object mental imagery induced by written text. *NeuroImage*, *120*, 309-322. doi:10.1016/j.neuroimage.2015.06.093

Andrews, M., Vigliocco, G., & Vinson, D. P. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463–498. doi:10.1037/a0016261.

Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology*, *96*, 43–55.

Anthony, J. L., Lonigan, C. J., Burgess, S. R., Driscoll, K., Phillips, B. M., & Cantor, B. G. (2002). Structure of preschool phonological sensitivity: Overlapping sensitivity to rhyme, words, syllables, and phonemes. *Journal of Experimental Child Psychology, 82*, 65–92.

Arata, M., Imai, M., Okuda, J., Okada, H., & Matsuda, T. (2010). Gesture in language: How sound symbolic words are processed in the brain. *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*, 1374-11379.

Arbuckle, S. A., Yokoi, A., Pruszynski, J. A., & Diedrichsen, J. (2018). Stability of representational geometry across a wide range of fMRI activity levels. *bioRxiv*. doi:10.1101/266585.

Aryani, A., Conrad, M., Schmidtke, D., & Jacobs, A. (2018). Why 'piss' is ruder than 'pee'? The role of sound in affective meaning making. *PLOS ONE, 13*(6), e0198430. doi:10.1371/journal.pone.0198430.

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., & Thierry, G. (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex, 63*, 196-205. doi:10.1016/j.cortex.2014.08.025.

Asher J.E., Aitken M.R., Farooqi N., Kurmani S., Baron-Cohen, S. (2006). Diagnosing and phenotyping visual synaesthesia: a preliminary evaluation of the revised test of genuineness (TOG-R). *Cortex*, *42*,137-146.

Asher, J. E., Lamb, J. A., Brocklebank, D., Cazier, J., Maestrini, E., Addis, L., … Monaco, A. P. (2009). A Whole-Genome Scan and Fine-Mapping Linkage Study of Auditory-Visual Synesthesia Reveals Evidence of Linkage to Chromosomes 2q24, 5q33, 6p12, and 12p12. *The American Journal of Human Genetics*, *84*(2), 279-285. doi:10.1016/j.ajhg.2009.01.012

Audacity Team (2012). Audacity v2.0.1 [Computer program]. Retrieved from http://audacity.sourceforge.net/ Audacity ® software is copyright © 1999-2014 Audacity Team.

Bargary G., Mitchell K.J. (2008). Synesthesia and cortical connectivity. *Trends in Neuroscience, 31*, 335–342.

Bankieris, K., & Simner, J. (2015). What is the link between synaesthesia and sound symbolism? *Cognition*, *136*, 186–195. doi: 10.1016/j.cognition.2014.11.013.

Baron-Cohen S., Burt, L., Smith-Laittan, F., Harrison, J., Bolton, P. (1996). Synaesthesia: prevalence and familiality. *Perception*, *25*, 1073-1079.

Baron-Cohen, S., Wyke, M.A., Binnie, C. (1987). Hearing words and seeing colours: an experimental investigation of a case of synaesthesia. *Perception*, *16*, 761-767.

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes, 8*(5–6), 513–562. doi:10.1080/01690960344000026.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*(1), 617-645. doi: 10.1146/annurev.psych.59.103006.093639.

Basso, D., Pavan, A., Ricciardi, E., Fagioli, S., Vecchi, T., Miniussi, C., & Pietrini, P. (2012). Touching motion: rTMS on the human middle temporal complex interferes with tactile speed perception. *Brain Topography*, *25*(4), 389-398. doi:10.1007/s10548-012-0223-4.

Beauchamp, M.S. (2005a). Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics*, 5, 93-114.

Beauchamp, M.S. (2005b). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Current Opinion in Neurobiology*, 15,145-153.

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *Journal of Neuroscience*, 30(7), 2414-2417. doi:10.1523/jneurosci.4865-09.2010.

Bedny, M., & Caramazza, A. (2011). Perception, action, and word meanings in the human brain: the case from action verbs. *Annuals of the New York Academy of Sciences,* 1224(1), 81-95. doi: 10.1111/j.1749-6632.2011.06013.x

Bedny, M., Pascual-Leone, A., Dodell-Feder, D., Fedorenko, E. & Saxe, R. (2011). Language processing in the occipital cortex of congenitally blind adults. *Proceedings of the National Academy of Science USA*, 108, 4429-4434.

Ben-Artzi, E. & Marks, L.E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics*, 57, 1151-1162.

Ben-Artzi, E. and Marks, L.E. (1999). Processing linguistic and perceptual dimensions of speech: interactions in speeded classification. *Journal of Experimental Psychology: Human Perception & Performance*, 25:579-595.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B Methodological, 57*(1), 289–300.

Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language, 80*(2), 290–311. doi:10.1353/lan.2004.0056.

Berlin, B. (1994). Evidence for pervasive synaesthetic sound symbolism in ethnozoological nomenclature. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 77–93). New York: Cambridge University Press.

Bernstein, I.H. and Edelstein, B.A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*, 87, 241-247.

Bhat, J., Miller, L.M., Pitt, M.A. & Shahin, A.J. (2015). Putative mechanisms mediating tolerance for audiovisual stimulus onset asynchrony. *Journal of Neurophysiology*, 113, 1437-1450.

Bierwisch, M. (1967). Some semantic universals of German adjectivals. *Foundations of Language*, *3*(1), 1–36.

Billock, V. A., & Havig, P. R. (2018). A simple power law governs many sensory amplifications and multisensory enhancements. *Scientific Reports*, *8*(1). doi:10.1038/s41598-018-25973-w.

Binder, J.R., Westbury, C.F., McKiernan, K.A., Possing, E.T. & Medler, D.A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*, 17, 905-917.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, *113*(39), 10818-10823. doi:10.1073/pnas.1605782113.

Blazej, L. J., & Cohen-Goldberg, A. M. (2016). Multicolored words: Uncovering the relationship between reading mechanisms and synesthesia. *Cortex*, *75*, 160-179. doi:10.1016/j.cortex.2015.11.017

Blazhenkova, O. & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology*, 23, 638–663.

Boersma, P. & Weenink, D. (2012). PRAAT: doing phonetics by computer. Accessed at http://www.praat.org/.

Bolognini, N., Olgiati, E., Rossetti, A., & Maravita, A. (2010). Enhancing multisensory spatial orienting by brain polarization of the parietal cortex. *European Journal of Neuroscience, 31*(10), 1800-1806. doi:10.1111/j.1460-9568.2010.07211.x.

Bolognini, N., Rossetti, A., Casati, C., Mancini, F., & Vallar, G. (2011). Neuromodulation of multisensory perception: A tDCS study of the sound-induced flash illusion. *Neuropsychologia*, *49*(2), 231-237. doi:10.1016/j.neuropsychologia.2010.11.015.

Bonath, B., Tyll, S., Budinger, E., Krauel, K., Hopf, J., & Noesselt, T. (2013). Task-demands and audio-visual stimulus configurations modulate neural activity in the human thalamus. *NeuroImage*, *66*, 110-118. doi:10.1016/j.neuroimage.2012.10.018.

Bosworth, R. G., & Emmorey, K. (2010). Effects of iconicity and semantic relatedness on lexical access in American Sign *Language. Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1573–1581.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433-436. doi:10.1163/156856897x00357.

Brand, J., Monaghan, P., & Walker, P. (2018). The Changing Role of Sound-Symbolism for Small Versus Large Vocabularies. *Cognitive Science*, *42*(Suppl Suppl 2), 578–590. doi: 10.1111/cogs.12565.

Brang, D., Hubbard, E.M., Coulson, S., Huang, M., Ramachandran, V.S. (2010). Magnetoencephalography reveals early activation of V4 in grapheme-color synesthesia. Neuroimage, 53, 268–274.

Brang, D., Rouw, R., Ramachandran, V., & Coulson, S. (2011). Similarly shaped letters evoke similar colors in grapheme–color synesthesia. *Neuropsychologia*, *49*(5), 1355-1358. doi:10.1016/j.neuropsychologia.2011.01.002

Brang, D., Williams, L. E., & Ramachandran, V. S. (2012). Grapheme-color synesthetes show enhanced crossmodal processing between auditory and visual modalities. *Cortex*, *48*(5), 630-637. doi:10.1016/j.cortex.2011.06.008.

Brang, D., Taich, Z. J., Hillyard, S. A., Grabowecky, M., & Ramachandran, V. (2013). Parietal connectivity mediates multisensory facilitation. NeuroImage, 78, 396-401. doi:10.1016/j.neuroimage.2013.04.047.

Brauchli, C., Elmer, S., Rogenmoser, L., Burkhard, A., & Jäncke, L. (2017). Top-down signal transmission and global hyperconnectivity in auditory-visual synesthesia: Evidence from a functional EEG resting-state study. *Human Brain Mapping*, *39*(1), 522-531. doi:10.1002/hbm.23861.

Bredin, H. (1996). Onomatopoeia as a figure and a linguistic principle. *New Literary History, 27*(3), 555–569. doi:10.1353/nlh.1996.0031.

Brett, M., Anton, J.L., Valabregue, R., & Poline, J.B. (2002). Region of interest analysis using an SPM toolbox [abstract] Presented at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan. *NeuroImage, (16)* 2 CD-ROM.

Brogaard, B. (2013). Seeing as a non-experiental mental state: The case from synesthesia and visual imagery. In R. Brown (Ed.), *Consciousness inside and out: Phenomenology, neuroscience, and the nature of experience* (pp. 377-394). Berlin, Germany: Springer Science & Business Media.

Brown, R. W., Black, A. H., & Horowitz, A. E. (1955). Phonetic symbolism in natural languages. *Journal of Abnormal and Social Psychology, 50*(3), 388–393. doi:10.1037/h0046820.

Brunetti, R., Indraccolo, A., Mastroberardino, S., Spence, C. & Santangelo, V. (2017). The impact of cross-modal correspondences on working memory performance. *Journal of Experimental Psychology: Human Perception & Performance*, 43:819-831.

Cabrera, Á., & Billman, D. (1996). Language-driven concept learning: Deciphering Jabberwocky. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(2), 539-555. doi:10.1037//0278-7393.22.2.539.

Cabrera, D. & Morimoto, M. (2007). Influence of fundamental frequency and source elevation on the vertical localization of complex tones and complex tone pairs. *Journal of the Acoustical Society of America*, 122:478-488.

Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex, 11*(12), 1110-1123. doi:10.1093/cercor/11.12.1110.

Calvert, G. A., & Campbell, R. (2003). Reading speech from still and moving faces: The neural substrates of visible speech. *Journal of Cognitive Neuroscience, 15*(1), 57-70. doi:10.1162/089892903321107828.

Cappe, C., Thelen, A., Romei, V., Thut, G., & Murray, M. M. (2012). Looming signals reveal synergistic principles of multisensory integration. *Journal of Neuroscience*, *32*(4), 1171-1182. doi:10.1523/jneurosci.5517-11.2012.

Carling, G., & Johansson, N. (2014). Motivated language change: processes involved in the growth and conventionalization of onomatopoeia and sound symbolism. *Acta Linguistica Hafniensia*, *46*(2), 199-217. doi:10.1080/03740463.2014.990293.

Carmichael, D., Down, M., Shillcock, R., Eagleman, D., & Simner, J. (2015). Validating a standardised test battery for synesthesia: Does the Synesthesia Battery reliably detect synesthesia? *Consciousness and Cognition*, *33*, 375-385. doi:10.1016/j.concog.2015.02.001.

Carroll, L. (2016). *Through the looking-glass and what Alice found there*. Buffalo, NY: Author's Republic.

Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language, 30*(3), 348–369. doi:10.1016/0749-596X(91)90041-H.

Catricalà, M. & Guidi, A. (2015). Onomatopoeias: a new perspective around space, image schemas, and phoneme clusters. *Cognitive Processing*, 16(Suppl 1):S175-S178.

Cavanna, A.E. & Trimble, M.R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*, 129:564-583.

Cernak, M., Asaei, A., Honnet, P., Garner, P. N., & Bourlard, H. (2016). Sound pattern matching for automatic prosodic event detection. *Interspeech 2016*. doi:10.21437/interspeech.2016-875.

Chambers, C. D., Payne, J. M., & Mattingley, J. B. (2007). Parietal disruption impairs reflexive spatial attention within and between sensory modalities. *Neuropsychologia, 45*(8), 1715-1724. doi:10.1016/j.neuropsychologia.2007.01.001

Chambers, C. D., Stokes, M. G., & Mattingley, J. B. (2004). Modality-specific control of strategic spatial attention in parietal cortex. *Neuron*, *44*(6), 925-930. doi:10.1016/j.neuron.2004.12.009.

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, *13*(11), 1428-1432. doi:10.1038/nn.2641.

Chen, Y.-C., Huang, P.-C., Woods, A. & Spence, C. (2016a). When “bouba” equals “kiki”: cultural commonalities and cultural differences in sound-shape correspondences. *Scientific Reports*, 6:26681, doi:10.1038/srep26681.

Chen, T., Michels, L., Supekar, K., Kochalka, J., Ryali, S. & Menon, V. (2015). Role of the anterior insula cortex in integrative causal signaling during multisensory auditory-visual attention. *European Journal of Neuroscience*, 41:264-274.

Chen, L., Zhang, M., Ai, F., Xie, W., & Meng, X. (2016b). Crossmodal synesthetic congruency improves visual timing in dyslexic children. *Research in Developmental Disabilities*, *55*, 14-26. doi:10.1016/j.ridd.2016.03.010

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(05), 489-558. doi: 10.1017/S0140525X08004998.

Colizoli, O., Murre, J. M., Scholte, H. S., & Rouw, R. (2017). Creating Colored Letters: Familial Markers of Grapheme–Color Synesthesia in Parietal Lobe Activation and Structure. *Journal of Cognitive Neuroscience*, *29*(7), 1239-1252. doi:10.1162/jocn_a_01105.

Colonius, H., & Diederich, A. (2017). Measuring multisensory integration: from reaction times to spike counts. *Scientific Reports*, *7*(1). doi:10.1038/s41598-017-03219-5.

Connor, C. E., Brincat, S. L., & Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Current Opinion in Neurobiology*, *17*(2), 140-147. doi:10.1016/j.conb.2007.03.002.

Corrigan, R. (2007). An experimental analysis of the affective dimensions of deep vocabulary knowledge used in inferring the meaning of words in context. *Applied Linguistics, 28*(2), 211–240. doi:10.1093/applin/amm009.

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, and Computers, 36*(3), 384–387. doi:10.3758/BF03195585.

Crottaz-Herbette, S., & Menon, V. (2006). Where and when the anterior cingulate cortex modulates attentional response: combined fMRI and ERP evidence. *Journal of Cognitive Neuroscience,* 18:766–80.

Cytowic, R.E. (1989). *Synaesthesia: A union of the senses*. New York, NY: Springer.

Cytowic, R. E. (2003). *The man who tasted shapes*. Cambridge, MA: Thorverton.

Cuskley, C., & Kirby, S. (2013). Synesthesia, Cross-Modality, and Language Evolution. *Oxford Handbooks Online*. doi:10.1093/oxfordhb/9780199603329.013.0043

Cuskley, C., Simner, J., & Kirby, S. (2015). Phonological and orthographic influences in the bouba–kiki effect. *Psychological Research*, *81*(1), 119-130. doi:10.1007/s00426-015-0709-2.

Day, S. (2005). Some demographic and socio-cultural aspects of synesthesia. In L. C. Robertson & N. Sagiv (Eds.), *Synesthesia: Perspectives from cognitive neuroscience* (pp. 11-33). Oxford University Press.

De Deyne, S., Peirsman, Y., & Storms, G. (2009). Sources of semantic similarity. In N. Taatgen & H. vanRijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1384–1389). Mahwah, NJ: Lawrence Erlbaum.

DeFife, L. C., Nygaard, L. C., & Namy, L. L. (2018). Cross-linguistic consistency and within-language variability of sound symbolism in nature languages. Manuscript in preparation.

Deroy, O., & Spence, C. (2013). Are we all born synaesthetic? Examining the neonatal synaesthesia hypothesis. *Neuroscience & Biobehavioral Reviews*, *37*(7), 1240-1253. doi:10.1016/j.neubiorev.2013.04.001.

Deroy, O., & Spence, C. (2013). Why we are not all synesthetes (not even weakly so). Psychonomic Bulletin & Review, 20(4), 643-664. doi:10.3758/s13423-013-0387-2.

Deroy, O., & Spence, C. (2016). Lessons of synaesthesia for consciousness: Learning from the exception, rather than the general. *Neuropsychologia*, *88*, 49-57. doi:10.1016/j.neuropsychologia.2015.08.005.

de Saussure, F. D. (1959). Chapter 2. In *Course in general linguistics*, (pp. 76-89). New York: Philosophical Library.

de Saussure, F.D. (2011). General principles: Nature of the linguistic sign. In P. Meisel & H. Saussy (Eds.), *Course in general linguistics* (W. Baskin, Trans.) (pp. 65-70). New York, NY: Columbia University Press.

Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *Journal of Neuroscience*, *33*(48), 18906-18916. doi:10.1523/jneurosci.3809-13.2013.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*(781). doi:10.3389/fpsyg.2014.00781.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences,19*(10), 603-615. doi:10.1016/j.tics.2015.07.013.

Dixon, M., Smilek, D., Duffy, P., Zanna, M., & Merikle, P. (2006). The role of meaning in grapheme-colour synaesthesia. *Cortex, 42*(2), 243-252. doi:10.1016/s0010-9452(08)70349-6.

Dixon, M. J., Smilek, D., & Merikle, P. M. (2004). Not all synaesthetes are created equal: Projector versus associator synaesthetes. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(3), 335-343. doi:10.3758/cabn.4.3.335.

Dojat, M., Pizzagalli, F., & Hupé, J.-M. (2018). Magnetic resonance imaging does not reveal structural alterations in the brain of grapheme-color synesthetes. PLoS ONE, 13(4), e0194422. doi: 10.1371/journal.pone.0194422.

Dolscheid, S., Hunnius, S., Casasanto, D. and Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychological Science*, 25:1256-1261.

Dovern, A., Fink, G. R., Fromme, A. C., Wohlschlager, A. M., Weiss, P. H., & Riedl, V. (2012). Intrinsic network connectivity reflects consistency of synesthetic experiences. *Journal of Neuroscience*, *32*(22), 7614-7621. doi:10.1523/jneurosci.5401-11.2012.

Downar, J., Crawley, A.P., Mikulis, D.J. & Davis, K.D. (2002). A cortical network sensitive to stimulus salience in a neural behavioral context across multiple sensory modalities. *Journal of Neurophysiology*, 87:615-620.

Drijvers, L., Zaadnoordijk, L., & Dingemanse, M. (2015). Sound-symbolism is disrupted in dyslexia: Implications for the role of cross-modal abstraction processes. In D. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, et al. (Eds.), Proceedings of the 37th Annual Meeting of the Cognitive Science Society (CogSci 2015), pp. 602-607. Austin, Tx: Cognitive Science Society.

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on ‘sensory-specific’ brain regions, neural responses, and judgments. *Neuron*, *57*(1), 11-23. doi:10.1016/j.neuron.2007.12.013.

Duncan, J. (2013). The structure of cognition: attentional episodes in mind and brain. *Neuron*, 80:35-50.

Duvernoy, H.M. (1999). *The Human Brain. Surface, Blood Supply and Three-dimensional Sectional Anatomy*. New York: Springer.

Eagleman, D.M., Kagan, A.D., Nelson, S.S., Sagaram, D., & Sarma, A.K. (2007) A standardized test battery for the study of Synesthesia. *Journal of Neuroscience Methods*, *159*(1):139-145. doi:10.1016/j.jneumeth.2006.07.012.

Eitan, Z. & Timmers, R. (2010). Beethoven’s last piano sonata and those who follow crocodiles: cross-domain mappings of auditory pitch in a musical context. *Cognition*, 114:405-422.

Eklund, A., Nichols, T.E. & Knutsson, H. (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences USA*, 113:7900-7905.

Erickson, L.C., Heeg, E., Rauschecker, J.P. & Turkeltaub, P.E. (2014). An ALE meta-analysis on the audiovisual integration of speech signals. *Human Brain Mapping*, 35:5587-5605.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429-433. doi:10.1038/415429a.

Evans, A., Collins, D., Mills, S., Brown, E., Kelly, R., & Peters, T. (1993). 3D statistical neuroanatomical models from 305 MRI volumes. *1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, 1813-1817. doi:10.1109/nssmic.1993.373602.

Evans, K.K. and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, 10:6 doi: 10.1167/10.1.6.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *The Journal of Neuroscience*, *22*(13), 5749-5759. doi:10.1523/jneurosci.22-13-05749.2002.

Farina, F. R., Mitchell, K. J., & Roche, R. A. (2017). Synaesthesia lost and found: Two cases of person- and music-colour synaesthesia. *European Journal of Neuroscience*, *45*(3), 472-477. doi:10.1111/ejn.13492.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). From the cover: Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences*, *103*(32), 12203-12208.

Fedorenko, E., Behr, M. K., & Kanwisher, N. G. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences USA*, 108:16428-16433.

Fedorenko, E., Duncan, J. & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences USA*, 110:16616-16621.

Fedorenko, E., Hsieh, P.J., Nieo-Castañón, A., Whitfield-Gabrieli, S. & Kanwisher, N. (2010). New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *Journal of Neurophysiology*, 104:1177-1194.

Fernández, L.M., Visser, M., Ventura-Campos, N., Ávila, C. & Soto-Faraco, S. (2015). Top-down attention regulates the neural expression of audiovisual integration. *NeuroImage,* 116:272-285.

Ferrand, C. T. (2002). Harmonics-to-noise ratio: An index of vocal aging. *Journal of Voice*, *16*(4), 480-487. doi:10.1016/s0892-1997(02)00123-6.

Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience*, *14*(6), 429-442. doi:10.1038/nrn3503.

Fitch, W.T. (2000). The evolution of speech: a comparative review. *Trends Cognitive Sciences*, 4, 258-267.

Flournoy, T. (1893). *Des phenomenes de synopsie*. Paris: Felix Alcan.

Flumini, A., Ranzini, M., & Borghi, A. M. (2014). Sound–shape correspondences with everyday objects figures. *Journal of Memory and Language*, *76*, 47–60. doi:10.1016/j.jml.2014.06.004.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A. et al. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magnetic Resonance in Medicine*, 33:636-647.

Fort, M., Martin, A. & Peperkamp, S. (2014). Consonants are more important than vowels in the Bouba-Kiki effect. *Language & Speech*, 58:247-266.

Fulbright, R.K., Manson, S.C., Skudlarski, P., Lacadie, C.M., & Gore, J.C. (2003). Quantity determination and the distance effect with letters, numbers, and shapes: A functional MR imaging study of number processing. *American Journal of Neuroradiology*, 24(2): 193-200.

Gallace, A. and Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, 68:1191-1203.

Galton, F. (1880). Visualised numerals. *Nature*, *21*(536), 323-323. doi:10.1038/021323a0.

Ganis, G., Thompson, W.L. & Kosslyn, S.M. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cognitive Brain Research*, 20:226-241.

Gasser, M. (2004). The origins of arbitrariness in language. *Proceedings of the Cognitive Science Society Conference*. (pp.434-439). Hillsdale, NJ: LEA.

Gasser, M., Sethuraman, N., & Hockema, S. (2005). Iconicity in expressives: An empirical investigation. In S. Rice & J. Newman (Eds.), *Experimental and empirical methods*. Stanford, CA: CSLI Publications.

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28,* 887–898.

Gebels, G. (1969). An investigation of phonetic symbolism in different cultures. Journal of *Verbal Learning and Verbal Behavior, 8*(2), 310–312. doi:10.1016/S0022-5371(69)80083-6.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, *15*(4), 870-878. doi:10.1006/nimg.2001.1037.

Gentilucci, M., & Corballis, M. C. (2006). From manual gesture to speech: A gradual transition. *Neuroscience and Biobehavioral Reviews, 30*(7), 949–960. doi:10.1016/j.neubiorev.2006.02.004.

Getz, L. M., & Kubovy, M. (2018). Questioning the automaticity of audiovisual correspondences. *Cognition*, *175*, 101-108. doi:10.1016/j.cognition.2018.02.015.

Ghazanfar, A., & Schroeder, C. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278-285. doi:10.1016/j.tics.2006.04.008.

Gibbs, R. W. (2003). Embodied experience and linguistic meaning. *Brain and Language, 84*(1), 1–15. doi:10.1016/S0093-934X(02)00517-5.

Gilron, R., Rosenblatt, J., Koyejo, O., Poldrack, R. A., & Mukamel, R. (2017). What's in a pattern? Examining the type of signal multivariate analysis uncovers at the group level. *NeuroImage*, *146*, 113-120. doi:10.1016/j.neuroimage.2016.11.019.

Girard, S., Collignon, O., & Lepore, F. (2010). Multisensory gain within and across hemispaces in simple and choice reaction time paradigms. *Experimental Brain Research*, *214*(1), 1-8. doi:10.1007/s00221-010-2515-9.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, *43*(3), 379–401. doi:10.1006/jmla.2000.2714.

Glezer, L.S., Eden, G., Jiang, X., Luetje, M., Naopliello, E., Kim, J. & Riesenhuber, M. (2016). Uncovering phonological and orthographic selectivity across the reading network using fMRI-RA. *NeuroImage*, 138:248-256.

Goodhew, S. C., Freire, M. R., & Edwards, M. (2015). Enhanced semantic priming in synesthetes independent of sensory binding. *Consciousness and Cognition*, *33*, 443-456. doi:10.1016/j.concog.2015.02.019

Goodhew, S. C., & Kidd, E. (2017). Language use statistics and prototypical grapheme colours predict synaesthetes' and non-synaesthetes' word-colour associations. *Acta Psychologica*, *173*, 73-86. doi:10.1016/j.actpsy.2016.12.008.

Goswami, U. (1993). Phonological skills and learning to read. *Annals of the New York Academy of Sciences*, 682:296-311.

Goswami, U. (2008). The development of reading across languages. *Annals of the New York Academy of Sciences*, 1145:1-12.

Goubet, N., Durand, K., Schaal, B., & McCall, D. D. (2018). Seeing odors in color: Cross-modal associations in children and adults from two cultural environments. *Journal of Experimental Child Psychology*, *166*, 380-399. doi:10.1016/j.jecp.2017.09.007

Greenwald, A.G., McGhee, D.E. & Schwarz, J.L.K. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality & Social Psychology*, 74:1464-1480.

Hackett, T. A., De La Mothe, L. A., Ulbert, I., Karmos, G., Smiley, J., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *The Journal of Comparative Neurology*, *502*(6), 924-952. doi:10.1002/cne.21326.

Hamada, D., Yamamoto, H., & Saiki, J. (2017). Multilevel analysis of individual differences in regularities of grapheme–color associations in synesthesia. *Consciousness and Cognition*, *53*, 122-135. doi:10.1016/j.concog.2017.05.007

Hamano, S. (1998). *The sound-symbolic system of Japanese*. Stanford, CA: CSLI Publications.

Hamilton-Fletcher, G., Pisanski, K., Reby, D., Stefańczyk, M., Ward, J., & Sorokowska, A. (2018). The role of visual experience in the emergence of cross-modal correspondences. *Cognition, 175,* 114-121. doi:10.1016/j.cognition.2018.02.023.

Han, H., & Glenn, A. L. (2017). Evaluating methods of correcting for multiple comparisons implemented in SPM12 in social neuroscience fMRI studies: An example from moral psychology. *Social Neuroscience*, *13*(3), 257-267. doi:10.1080/17470919.2017.1324521.

Harrar, V., Aubin, S., Chebat, D., Kupers, R., & Ptito, M. (2017). The multisensory blind brain. *Mobility of Visually Impaired People,* 111-136. doi:10.1007/978-3-319-54446-5_4.

Hartwigsen, G., Baumgaertner, A., Price, C.J., Koehnke, M., Ulmer, S. & Siebner, H.R. (2010). Phonological decisions require both left and right supramarginal gyri. *Proceedings of the National Academy of Sciences USA*, 107:16494-16499.

Hasson, G., & Willems, F. (2018). Grounding the neurobiology of language in first principles: The necessity of non-language-centric explanations for language comprehension. *Cognition, 180,* 135-157. doi:10.1016/j.cognition.2018.06.018.

Haxby, J. V. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430. doi:10.1126/science.1063736.

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *NeuroImage, 62*(2), 852-855. doi:10.1016/j.neuroimage.2012.03.016.

Haynes, J., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience, 7*(7), 523-534. doi:10.1038/nrn1931.

Hebart, M. N.*, Görgen, K.*, & Haynes, J. (2015). The Decoding Toolbox (TDT): A versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, *8*. doi:10.3389/fninf.2014.00088. *equal contribution

Hein, G., Doehrmann, O., Müller, N.G., Kaiser, J., Muckli, L. Naumer, M.J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *Journal of Neuroscience*, 27:7881-7887.

Herold, D. S., Nygaard, L. C., Chicos, K. A., & Namy, L. L. (2011). The developing role of prosody in novel word interpretation. *Journal of Experimental Child Psychology, 108,* 229–241.

Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science*, *15*(5), 688-696. doi:10.1111/j.1467-7687.2012.01171.x.

Hinton, L., Nichols, J., & Ohala, J. (Eds.) (1994). *Sound Symbolism*. Cambridge, UK: Cambridge University Press.

Hockett, C. F. (1977). *The view from language: Selected essays 1948–1974.* Athens: The University of Georgia Press.

Hollien, H., Girard, G. T., & Coleman, R. F. (1977). Vocal fold vibratory patterns of pulse register phonation. *Folia Phoniatrica et Logopaedica*, *29*(3), 200-205. doi:10.1159/000264089.

Holmes, N. P., & Spence, C. (2005). Multisensory integration: Space, time and superadditivity. *Current Biology, 15*(18), R762-R764. doi:10.1016/j.cub.2005.08.058.

Hubbard, E. M., Brang, D., & Ramachandran, V. S. (2011). The cross-activation theory at 10. *Journal of Neuropsychology*, *5*(2), 152-177. doi:10.1111/j.1748-6653.2011.02014.x

Hubbard, E. M., & Ramachandran, V. (2005). Neurocognitive mechanisms of synesthesia. *Neuron, 48*(3), 509-520. doi: 10.1016/j.neuron.2005.10.012.

Hupé, J., & Dojat, M. (2015). A critical review of the neuroimaging literature on synesthesia. *Frontiers in Human Neuroscience, 9*. doi:10.3389/fnhum.2015.00103.

Hutzler, F. (2014). Reverse inference is not a fallacy per se: cognitive processes can be inferred from functional imaging data. *NeuroImage*, 84:1061-1069.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130298. doi: 10.1098/rstb.2013.0298

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109,* 54-65.

Imai, M., Miyazaki, M., Yeung, H. H., Hidaka, S., Kantartzis, K., Okada, H., & Kita, S. (2015). Sound symbolism facilitates word learning in 14-month-olds. *PLoS ONE, 10*(2), e0116494.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press: Oxford, UK.

Jamal, Y., Lacey, S., Nygaard, L. & Sathian, K. (2017). Interactions between auditory elevation, auditory pitch, and visual elevation during multisensory perception. *Multisensory Research*, 30:287-306.

James, T., & Stevenson, R. (2011). The use of fMRI to assess multisensory integration. *Frontiers in Neuroscience*, 131-146. doi:10.1201/b11092-11.

James, T.W., Stevenson, R.A., Kim, S., VanDerKlok, R.M. & James, K.H. (2011). Shape from sound: evidence for a shape operator in the lateral occipital complex. *Neuropsychologia*, 49:1807-1815.

Johnson, M. A., & Goldberg, A. E. (2013). Evidence for automatic accessing of constructional meaning: Jabberwocky sentences prime associated verbs. *Language and Cognitive Processes*, *28*(10), 1439-1452. doi:10.1080/01690965.2012.717632.

Jonas, C., Spiller, M. J., & Hibbard, P. (2017). Summation of visual attributes in auditory–visual crossmodal correspondences. *Psychonomic Bulletin & Review*, *24*(4), 1104-1112. doi:10.3758/s13423-016-1215-2.

Joseph, J. E. (2015). Iconicity in Saussure's Linguistic Work, and why it does not contradict the arbitrariness of the sign. *Historiographia Linguistica International Journal for the History of the Language Sciences*, *42*(1), 85-105. doi:10.1075/hl.42.1.05jos.

Kamitani, Y., & Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. *Current Biology*, *16*(11), 1096-1102. doi:10.1016/j.cub.2006.04.003.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679-685. doi:10.1038/nn1444.

Kamke, M. R., Vieth, H. E., Cottrell, D., & Mattingley, J. B. (2012). Parietal disruption alters audiovisual binding in the sound-induced flash illusion. *NeuroImage*, *62*(3), 1334-1341. doi:10.1016/j.neuroimage.2012.05.063.

Kanero, J., Imai, M., Okuda, J., Okada, H., & Matsuda, T. (2014). How Sound Symbolism Is Processed in the Brain: A Study on Japanese Mimetic Words. *PLoS ONE, 9*(5), e97905. doi:10.1371/journal.pone.0097905

Kang M-J, Kim Y, Shin J-Y, Kim C-Y. Graphemes Sharing Phonetic Features Tend to Induce Similar Synesthetic Colors. *Frontiers in Psychology*. 2017;8:337. doi:10.3389/fpsyg.2017.00337.

Kantartzis, K., Imai, M., and Kita, S. (2011). Japanese sound-symbolism facilitates verb learning in English speaking children. *Cognitive Science, 35*, 575–586.

Kay, C. L., Carmichael, D. A., Ruffell, H. E., & Simner, J. (2014). Colour fluctuations in grapheme-colour synaesthesia: The effect of clinical and non-clinical mood changes. *British Journal of Psychology, 106*(3), 487-504. doi:10.1111/bjop.12102.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, *99*(2), 349–364. doi:10.1037/0033-295X.99.2.349.

Kemmerer, D. (2009). How words capture visual experience: The perspective from Cognitive Neuroscience. In B. Malt & P. Wolff (Eds.), Words and the mind: How Words Capture Human Experience (pp. 287 – 327). New York: Oxford University press.

Kilian-Hutten, N., Valente, G., Vroomen, J., & Formisano, E. (2011). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *Journal of Neuroscience*, *31*(5), 1715-1720. doi:10.1523/jneurosci.4572-10.2011.

Kinnear, P. R. (2002). New Farnsworth-Munsell 100 hue test norms of normal observers for each year of age 5-22 and for age decades 30-70. *British Journal of Ophthalmology*, *86*(12), 1408-1411. doi:10.1136/bjo.86.12.1408.

Kirchner, H., Barbeau, E.J., Thorpe, S.J., Régis, J. & Liégeois-Chauvel, C. (2009). Ultra-rapid responses in the human frontal eye field region. *Journal of Neuroscience*, 29:7599-7606.

Kirschner, A., & Nikolić, D. (2017). One-shot synesthesia. *Translational Neuroscience*, *8*(1). doi:10.1515/tnsci-2017-0023.

Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics*, 35:379-415.

Kleiner M., Brainard D., Pelli D. (2007) "What's new in Psychtoolbox-3?" Perception 36 ECVP Abstract Supplement.

Kliper, R., Portuguese, S., & Weinshall, D. (2016). Prosodic analysis of speech and the underlying mental state. *Communications in Computer and Information Science*, 52-62. doi:10.1007/978-3-319-32270-4_6.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712-719. doi:10.1016/j.tins.2004.10.007.

Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports*, *7*, 5562. doi: 10.1038/s41598-017-05965-y.

Köhler, W. (1929). Gestalt Psychology. New York: Liveright Publishing Corporation.

Köhler, W. (1947). *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*. Liveright: New York, NY.

Koppensteiner, M., Stephan, P. & Jäschke, J.P. (2016). Shaking takete and flowing maluma: nonsense words are associated motion patterns. *PLoS ONE*, 11:e0150610, doi:10.1371/journal.pone.0150610

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition, 114(1)*, 19-28. doi: 10.1016/j.cognition.2009.08.016

Kragel, P. A., Carter, R. M., & Huettel, S. A. (2012). What makes a pattern? Matching decoding methods to data in multivariate pattern analysis. *Frontiers in Neuroscience*, *6*. doi:10.3389/fnins.2012.00162.

Kriegeskorte, N., & Bandettini, P. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. *NeuroImage*, *38*(4), 649-662. doi:10.1016/j.neuroimage.2007.02.022.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*. doi:10.3389/neuro.06.004.2008.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., … Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126-1141. doi:10.1016/j.neuron.2008.10.043.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, *12*(5), 535-540. doi:10.1038/nn.2303

Krugliak, A., & Noppeney, U. (2016). Synaesthetic interactions across vision and audition. *Neuropsychologia*, *88*, 65-73. doi:10.1016/j.neuropsychologia.2015.09.027

Kunihira, S. (1971). Effects of the expressive voice on phonetic symbolism. *Journal of Verbal Learning and Behavior*, 10:427-429.

Lacadie, C.M., Fulbright, R.K., Constable, R.T. & Papademetris, X. (2008). More accurate Talairach coordinates for neuroimaging using nonlinear registration. NeuroImage, 42:717-725. Accessed at http://sprout022.sprout.yale.edu/mni2tal/mni2tal.html

Lacey, S., Lin, J.B. & Sathian, K. (2011). Object and spatial imagery dimensions in visuo-haptic representations. *Experimental Brain Research*, 213(2-3), 267-273.

Lacey, S., Martinez, M., McCormick, K., & Sathian, K. (2016). Synesthesia strengthens sound symbolic crossmodal correspondences. *The European Journal of Neuroscience*, *44*(9), 2716–2721. doi: 10.1111/ejn.13381.

Lacey, S., Stilla, R., Deshpande, G., Zhao, S., Stephens, C., McCormick, K., … Sathian, K. (2017). Engagement of the left extrastriate body area during body-part metaphor comprehension. *Brain and Language, 166,* 1-18. doi:10.1016/j.bandl.2016.11.004.

Lacey, S., Stilla, R., & Sathian, K. (2012). Metaphorically feeling: Comprehending textural metaphors activates somatosensory cortex. *Brain and Language*, *120*(3), 416-421. doi:10.1016/j.bandl.2011.12.016.

Lacey, S., Tal, N., Amedi, A. & Sathian, K. (2009). A putative model of multisensory object representation. *Brain Topography*, 21:269-274.

Ladeira, A., Fregni, F., Campanhã, C., Valasek, C. A., De Ridder, D., Brunoni, A. R., & Boggio, P. S. (2011). Polarity-dependent transcranial direct current stimulation effects on central auditory processing. *PLoS ONE*, *6*(9), e25399. doi:10.1371/journal.pone.0025399.

Lakoff, G., & Johnson, M. (1980). Metaphors we live by. Chicago, IL: University of Chicago Press.

Lamichhane, B., Adhikari, B.M. & Dhamala, M. (2016). The activity in the anterior insulae is modulated by perceptual decision-making difficulty. *Neuroscience,* 327:79-94.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. In Psychological Review, 104(2), 211–240.

Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y., Field, A. S., & Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of Cognitive Neuroscience*, *14*(3), 420-429. doi:10.1162/089892902317361930.

Laurienti, P. J., Perrault, T. J., Stanford, T. R., Wallace, M. T., & Stein, B. E. (2005). On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Experimental Brain Research*, *166*(3-4), 289-297. doi:10.1007/s00221-005-2370-2.

Lee, H., & Noppeney, U. (2011a). Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proceedings of the National Academy of Sciences*, *108*(51), E1441-E1450. doi:10.1073/pnas.1115267108.

Lee, H., & Noppeney, U. (2011b). Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. *Journal of Neuroscience*, *31*(31), 11338-11350. doi:10.1523/jneurosci.6510-10.2011.

Lenhard, W. & Lenhard, A. (2016). Calculation of Effect Sizes. *Psychometrica*. doi: 10.13140/RG.2.1.3478.4245 Accessed at https://www.psychometrica.de/effect_size.html.

Lewis, R., & Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *Journal of Neuroscience*, *30*(37), 12329-12339. doi:10.1523/jneurosci.5745-09.2010.

Liberman, I., Shankweiler, D., Fischer, F., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201–212.

List, S.M., McCormick, K., Lacey, S., Sathian, K., & Nygaard, L.C. (under review a). Acoustic and visual stimulus parameters underlying sound symbolic crossmodal correspondences.

Journal of Cognitive Science. (*bioRxiv* Pre-print: https://www.biorxiv.org/content/early/2019/01/11/517581)

Locke, J. (1715). Book III On words: Chapter I Of words or language in general. In *An essay concerning human understanding*. London: Printed for J. Churchill; and Samuel Manship.

Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: a review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, *6*, 1246. doi: 10.3389/fpsyg.2015.01246

Lockwood, G., Dingemanse, M., & Hagoort, P. (2016). Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1274-1281. doi:10.1037/xlm0000235

Lockwood, G., Hagoort, P. & Dingemanse, M. (2016). How iconicity helps people learn new words: neural correlates and individual differences in sound-symbolic bootstrapping. *Collabra*, 2:7, doi: 10.1525/Collabra.42

Lourenco, S. F., & Longo, M. R. (2011). Origins and development of generalized magnitude representation. In S. Dehaene, & E. Brannon (Eds.), *Space, time and number in the brain: Searching for the foundations of mathematical thought* (pp. 225–244). London: Academic.

Lu, Y., & Cooke, M. (2009). The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. *Speech Communication*, *51*(12), 1253-1262. doi:10.1016/j.specom.2009.07.002.

Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *Proceedings of the National Academy of Sciences*, *108*(51), 20661-20665. doi:10.1073/pnas.1112605108

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrences. *Behavior Research Methods, Instruments, and Computers, 28*(2), 203–208. doi:10.3758/BF03204766.

Lupyan, G., & Casasanto, D. (2014). Meaningless words promote meaningful categorization. *Language and Cognition*, *7*(02), 167-193. doi:10.1017/langcog.2014.21.

Lupyan, G., & Spivey, M. J. (2010). Making the invisible visible: Verbal but not visual cues enhance visual detection (J. Lauwereyns, Ed.). *PLoS ONE*, 5(7), E11452. doi: 10.1371/journal.pone.0011452.

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. Proceedings of the National Academy of Sciences, 110(35), 14196-14201. doi: 10.1073/pnas.1303312110

Lupyan, G., & Winter, B. (2018). Language is more abstract than you think, or, why aren't languages more iconic? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1752), 20170137. doi:10.1098/rstb.2017.0137.

Lycan, W. (2006). Representational theories of consciousness. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.

Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge, UK: Cambridge University Press.

Lyons, J. (1970). *New horizons in linguistics*. Middlesex, UK: Penguin Books.

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends in Neurosciences*, *28*(5), 264-271. doi:10.1016/j.tins.2005.03.008.

Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. Cognitive Psychology, 15(2), 197-237. doi:10.1016/0010-0285(83)90009-9.

Marchant, J.L., Ruff, C.C. & Driver, J. (2012). Audiovisual asynchrony enhances BOLD responses in a brain network including multisensory STS while also enhancing target-detection performance for both modalities. *Human Brain Mapping*, 33:1212-1224.

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. Journal of Experimental Psychology: Human Perception and Performance, 13(3), 384.

Marks, L.E. (2011). Synaesthesia, now and then. *Intellectica*, 55, 47-80.

Marques, L. M., Lapenta, O. M., Merabet, L. B., Bolognini, N., & Boggio, P. S. (2014). Tuning and disrupting the brain-modulating the McGurk illusion with electrical stimulation. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00533.

Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J., Maeder, P. P., Clarke, S., & Meuli, R. A. (2006). Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cerebral Cortex*, *17*(7), 1672-1679. doi:10.1093/cercor/bhl077.

Mastroberardino, S., Santangelo, V. & Macaluso, E. (2015). Crossmodal semantic congruence can affect visuo-spatial processing and activity of the fronto-parietal attention networks. *Frontiers in Integrative Neuroscience*, 9:45, doi:10.3389/fnint.2015.00045.

Mathôt, S., Grainger, J., & Strijkers, K. (2017). Pupillary responses to words that convey a sense of brightness or darkness. *Psychological Science*, *28*(8), 1116-1124. doi:10.1177/0956797617702699.

Matusz, P. J., Wallace, M. T., & Murray, M. M. (2017). A multisensory perspective on object memory. *Neuropsychologia*, *105*, 243-252. doi:10.1016/j.neuropsychologia.2017.04.008.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, *9*(3), 316-322. doi: 10.1111/j.1467-7687.2006.00495.x.

Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., & Lancaster, J. (1995). A probabilistic atlas of the human brain: Theory and rationale for its development. *NeuroImage*, *2*(2), 89-101. doi:10.1006/nimg.1995.1012.

Mazzoni, D., & Dannenberg, R. (2000). *Audacity (software)*. Pittsburg, PA: Carnegie Mellon University.

McCormick, K., Kim, J.Y., List, S. & Nygaard, L.C. (2015). Sound to meaning mappings in the bouba-kiki effect. *Proceedings 37th Annual Meeting Cognitive Science Society*, 1565-1570.

McCormick, K., Lacey, S., Stilla, R., Nygaard, L. C., & Sathian, K. (2018). Neural basis of the crossmodal correspondence between auditory pitch and visuospatial elevation. *Neuropsychologia*, *112*, 19-30. doi:10.1016/j.neuropsychologia.2018.02.029.

McCormick, K., Lacey, S., Stilla, R., Nygaard, L.C., & Sathian, K.(2018). Neural basis of the sound-symbolic crossmodal correspondence between auditory pseudowords and visual shapes. *BioRxiv*.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748. doi:10.1038/264746a0.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, and Computers, 37*(4), 547–559. doi:10.3758/BF03192726.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology, 126*(2), 99–130. doi:10.1037/0096-3445.126.2.99.

Meier, R. P., Mauk, C. E., Cheek, A., & Moreland, C. J. (2008). The form of children's early signs: Iconic or motoric determinants? *Language Learning and Development, 4,* 63–98.

Melara, R. D., & Marks, L. E. (1990a). Dimensional interactions in language processing: investigating directions and levels of crosstalk. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(4), 539-554. doi: 10.1037//0278-7393.16.4.539

Melara, R. D., & Marks, L. E. (1990b). Processes underlying dimensional interactions: Correspondences between linguistic and nonlinguistic dimensions. *Memory & Cognition,* 18(5), 477-495. doi: 10.3758/BF03198481

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, 116(4), 323-336. doi: 10.1037/0096-3445.116.4.323.

Merabet, L. B., Hamilton, R., Schlaug, G., Swisher, J. D., Kiriakopoulos, E. T., Pitskel, N. B., … Pascual-Leone, A. (2008). Rapid and reversible recruitment of early visual cortex for touch. *PLoS ONE, 3*(8), e3046. doi:10.1371/journal.pone.0003046.

Merabet, L. B., Theoret, H., & Pascual-Leone, A. A. (2003). Transcranial magnetic stimulation as an investigative tool in the study of visual function. *Optometry and Vision Science, 80*(5), 356-368. doi:10.1097/00006324-200305000-00010.

Meredith, M., & Stein, B. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, *221*(4608), 389-391. doi:10.1126/science.6867718.

Meteyard, L., Stoppard, E., Snudden, D., Cappa, S. F., & Vigliocco, G. (2015). When semantics aids phonology: A processing advantage for iconic word forms in aphasia. *Neuropsychologia*, 76, 264-275. doi:10.1016/j.neuropsychologia.2015.01.042.

Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *NeuroImage, 53*(1), 103-118. doi:10.1016/j.neuroimage.2010.05.051.

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Visual-auditory multisensory object recognition in humans: A high-density electrophysiological study. *Cerebral Cortex, 14*(4), 452–465. doi:10.1093/cercor/bhh007.

Monaghan, P., & Christiansen, M. H. (2006). Why form-meaning mappings are not entirely arbitrary in language. *Cognitive Science Journal*, 1838-1843. Retrieved September 21, 2013, from http://csjarchive.cogsci.rpi.edu

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology, 55*(4), 259–305. doi:10.1016/j.cogpsych.2006.12.001.

Monaghan, P., Christiansen, M.H. & Fitneva, S.A. (2011). The Arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General* (140)3, 325-347. doi: 10.1037/a0022924

Monaghan, P., Mattock, K. & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 38:1152-1164.

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130299. doi: 10.1098/rstb.2013.0299.

Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch–object correspondences in young children. *Cognitive, Affective, and Behavioral Neuroscience, 4*(2),133–136. doi:10.3758/CABN.4.2.133.

Moore, B.C.J. (2012). Psychology of Hearing, 6th Edition. Leiden, The Netherlands: Brill.

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., & Sams, M. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *NeuroImage*, *30*(2), 563-569. doi:10.1016/j.neuroimage.2005.10.002.

Mozolic, J. L., Joyner, D., Hugenschmidt, C. E., Peiffer, A. M., Kraft, R. A., Maldjian, J. A., & Laurienti, P. J. (2008). Cross-modal deactivations during modality-specific selective attention. *BMC Neurology*, *8*(1). doi:10.1186/1471-2377-8-35.

Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—an introductory guide. *Social Cognitive and Affective Neuroscience*, *4*(1), 101-109. doi:10.1093/scan/nsn044.

Murray, M. M., Thelen, A., Ionta, S., & Wallace, M. T. (2018). Contributions of intraindividual and interindividual differences to multisensory processes. *Journal of Cognitive Neuroscience,* 1-17. doi:10.1162/jocn_a_01246.

Murray, M. M., Lewkowicz, D. J., Amedi, A., & Wallace, M. T. (2016). Multisensory processes: A balancing act across the lifespan. *Trends in Neurosciences, 39*(8), 567-579. doi:10.1016/j.tins.2016.05.003.

Namy, L. L. (2008). Recognition of iconicity doesn't come for free. *Developmental Science, 11,* 841–846.

Namy, L. L., & Nygaard, L. C. (2008). Perceptual-motor constraints on sound to meaning correspondence in language. *Behavioral and Brain Sciences, 31*(5), 528–529. doi:10.1017/S0140525X08005190.

Narsude, M., Gallichan, D., Van der Zwaag, W., Gruetter, R., & Marques, J. P. (2015). Three-dimensional echo planar imaging with controlled aliasing: A sequence for high temporal resolution functional MRI. *Magnetic Resonance in Medicine*, *75*(6), 2350-2361. doi:10.1002/mrm.25835.

Neufeld, J., Sinke, C., Zedler, M., Dillo, W., Emrich, H., Bleich, S., & Szycik, G. (2012). Disinhibited feedback as a cause of synesthesia: Evidence from a functional connectivity study on auditory-visual synesthetes. *Neuropsychologia*, *50*(7), 1471-1477. doi:10.1016/j.neuropsychologia.2012.02.032

Newell, F. N., & Mitchell, K. J. (2016). Multisensory integration and cross-modal learning in synaesthesia: A unifying model. *Neuropsychologia*, *88*, 140-150. doi:10.1016/j.neuropsychologia.2015.07.026.

Newmeyer, F. J. (1993). Iconicity and generative grammar. *Language, 68*(4), 756–796. doi:10.2307/416852.

Nielsen, A., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. Canadian Journal of

Experimental Psychology/Revue Canadienne De Psychologie Expérimentale, 65(2), 115-124. doi: 10.1037/a0022268.

Nielsen, A.K.S. (2016). *Systematicity, motivatedness, and the structure of the lexicon* (PhD thesis). University of Edinburgh.

Noesselt, T., Bergmann, D., Heinze, H.-J., Münte, T. & Spence, C. (2012). Coding of multisensory temporal patterns in human superior temporal sulcus. *Frontiers in Integrative Neuroscience*, 6:64, doi: 10.3389/fnint.2012.00064

Noppeney, U. (2012). Characterization of multisensory integration with fMRI. In M.M. Murray & M.T. Wallace (Eds.), *The Neural Bases of Multisensory Processes*. (pp 233-252). CRC Press.

Noppeney, U., Josephs, O., Hocking, J., Price, C.J. & Friston, K.J. (2008). The effect of prior visual information on recognition of speech and sounds. *Cerebral Cortex*, 18:598-609.

Nordmark, P. F., Pruszynski, J. A., & Johansson, R. S. (2012). BOLD responses to tactile stimuli in visual and auditory cortex depend on the frequency content of stimulation. *Journal of Cognitive Neuroscience*, *24*(10), 2120-2134. doi:10.1162/jocn_a_00261.

Novack, M. A., Goldin-Meadow, S., & Woodward, A. L. (2015). Learning from gesture: How early does it happen? *Cognition, 142*, 138–147.

Nuckolls, J. B. (1999). The case for sound symbolism. *Annual Review of Anthropology,28*(1), 225-252. doi: 10.1146/annurev.anthro.28.1.225

Nye, E. (2000). Linguistic and poetic sound symbolism. In *Literary and Linguistic Theories in Eighteenth-Century France.* (pp 129-178). Oxford University Press. doi:10.1093/acprof:oso/9780198160120.003.0005.

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, *112*(1), 181-186. doi: 10.1016/j.cognition.2009.04.001.

Oberhuber, M., Hope, T.M.H., Seghier, M.L., Jones, O.P., Prejawa, S., Green, D.W. & Price, C.J. (2016). Four functionally distinct regions in the left supramarginal gyrus support word processing. *Cerebral Cortex*, 26:4212-4226.

Optseq2 [Computer software]. (2006). Retrieved from https://surfer.nmr.mgh.harvard.edu/optseq/.

Osgood, C. E. (1969). On the whys and wherefores of E. P. A. *Journal of Personality and Social Psychology*, *12*(3), 194–199. doi:10.1037/h0027715.

Ozcaliskan, S., & Goldin-Meadow, S. (2011). Is there an iconic gesture spurt at 26 months? In G. Stam & M. Ishino (Eds.), *Integrating gestures: The interdisciplinary nature of gesture* (pp. 163–174). Amsterdam: John Benjamins.

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(2), 173-186. doi:10.1016/j.jecp.2012.05.004.

Paffen, C. L., Van der Smagt, M. J., & Nijboer, T. C. (2015). Cross-modal, bidirectional priming in grapheme-color synesthesia. *Consciousness and Cognition*, *33*, 325-333. doi:10.1016/j.concog.2015.01.020.Pajula, J., & Tohka, J. (2014). Effects of spatial smoothing on inter-subject correlation based analysis of FMRI. *Magnetic Resonance Imaging*, *32*(9), 1114-1124. doi:10.1016/j.mri.2014.06.001.

Palmeri, T. J., Blake, R., Marois, R., Flanery, M. A., & Whetsell, W. (2002). The perceptual reality of synesthetic colors. *Proceedings of the National Academy of Sciences*, *99*(6), 4127-4131. doi:10.1073/pnas.022049399.

Paradis, C., & Willners, C. (2011). Antonymy: From convention to meaning-making. *Review of Cognitive Linguistics, 9*(2), 367–391. doi:10.1075/rcl.9.2.02par.

Paradis, C., Willners, C., & Jones, S. (2009). Good and bad opposites using textual and experimental techniques to measure antonym canonicity. *The Mental Lexicon, 4*(3), 380–429. doi:0.1075/ml.4.3.04par.

Parault, S. J., & Parkinson, M. M. (2008). Sound symbolic word learning in the middle grades. *Contemporary Educational Psychology, 33*(4), 647–671. doi:10.1016/j.cedpsych.2007.07.003.

Parault, S. J., & Schwanenflugel, P. J. (2006). Sound symbolism: A possible piece in the puzzle of word learning. *Journal of Psycholinguistic Research, 35,* 329–351. doi:10.1007/s10936-006-9018-7.

Parise, C.V., Knorre, K. and Ernst, M.O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences USA*, 111:6104-6108.

Parise, C.V., & Pavani, F. (2011). Evidence of sound symbolism in simple vocalizations. Experimental Brain Research, 214(3), 373-80.

Parise, C. V., & Spence, C. (2009). 'When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes . *PLoS ONE, 4(5)*, E5664. doi: 10.1371/journal.pone.0005664

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research*, 220(3-4), 319-333. doi: 10.1007/s00221-012-3140-6.

Parkinson, C., Kohler, P.J., Sievers, B. and Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: evidence from a remote population. *Perception*, 41:854-861.

Pasalar, S., Ro, T., & Beauchamp, M. S. (2010). TMS of posterior parietal cortex disrupts visual tactile multisensory integration. *European Journal of Neuroscience*, *31*(10), 1783-1790. doi:10.1111/j.1460-9568.2010.07193.x.

Pascual-Leone, A. & Hamilton, R.H. (2001). The metamodal organization of the brain. *Progress in Brain Research,* 134:427-445.

Peiffer-Smadja, N. (2010). *Exploring the bouba/kiki effect: A behavioral and fMRI study* (Unpublished master's thesis). Université Paris Descartes.

Peiffer-Smadja, N., & Cohen, L. (2019). The cerebral bases of the bouba-kiki-effect. NeuroImage, 186, 679-689.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437-442. doi:10.1163/156856897x00366.

Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology, 1*, 1–15. doi:10.3389/fpsyg.2010.00227.

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651), 20130300. doi:10.1098/rstb.2013.0300.

Perrodin, C., Kayser, C., Logothetis, N. K., & Petkov, C. I. (2014). Auditory and visual modulation of temporal lobe neurons in voice-sensitive and association cortices. *Journal of Neuroscience, 34*(7), 2524-2537. doi:10.1523/jneurosci.2805-13.2014.

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS ONE, 10*(9), e0137147.

Petrou M., & Petrou C. (2011). Image transformations. *Image processing: The fundamentals*, 47-176. doi:10.1002/9781119994398.ch2.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., & Spivey, M. J. (2013). Computational grounded cognition: A new alliance between grounded cognition and computational modeling. Frontiers in Psychology, 3(612), 1-11. doi: 10.3389/fpsyg.2012.00612

Piazza, M., Izard, V., Pinel, P., Le Bihan, D. & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, 44:547-555.

Piazza, M., Pinel, P., Le Bihan, D. & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, 53:293-305.

Pinel, P., Piazza, M., Le Bihan, D. & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron*, 41:983-993.

Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. Harper Collins.

Pitcher, D., Garrido, L., Walsh, V., & Duchaine, B. C. (2008). Transcranial magnetic stimulation disrupts the perception and embodiment of facial expressions. *Journal of Neuroscience, 28*(36), 8929-8933. doi:10.1523/jneurosci.1450-08.2008.

Plato (360 B.C.E.). Cratylus by Plato. Ed. Jowett, B. Retrieved June 20, 2018, from http://classics.mit.edu/Plato/cratylus.html.

Poldrack, R. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10:59-63.

Poldrack, R. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron*, 72:692-697.

Popp, M., Trumpp, N. M., & Kiefer, M. (2016). Feature-specific event-related potential effects to action- and sound-related verbs during visual word recognition. *Frontiers in Human Neuroscience*, *10*. doi:10.3389/fnhum.2016.00637.

Powell, M. (2012). A meditation partially composed in a D.C. coffeehouse because there isn't anything better to do in this city of dead white fathers... In Q. Driskill, D. H. Justice, D. Miranda, L. Tatonetti, I. Allegra, & L. E. Cruz (Authors), Sovereign erotics a collection of two-spirit literature (pp. 84-85). Tucson: University of Arizona Press.

Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *Journal of Neuroscience*, *29*(39), 12265-12274. doi:10.1523/jneurosci.3501-09.2009.

Pratt, C.C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, 13:278-285.

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, *1191*(1), 62-88. doi:10.1111/j.1749-6632.2010.05444.x.

Price, C.J., Moore, C.J., Humphreys, G.W. & Wise, R.J.S. (1997). Segregating semantic from phonological processes during reading. *Journal of Cognitive Neuroscience*, 9:727-733.

Price, C.J., Wise, R.J.S & Frackowiak, R.S.J. (1996). Demonstrating the implicit processing of visually presented words and pseudowords. *Cerebral Cortex*, 6:62-70.

Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. Science, 275(5306), 1604-1610. doi:10.1126/science.275.5306.1604.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, *103*(20), 7865-7870. doi:10.1073/pnas.0509989103.

Quak, M., London, R.E & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in Human Neuroscience*, 9:197. doi: 10.3389/fnhum.2015.00197.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Raczkowski, D., Kalat, J. W., & Nebes, R. (1974). Reliability and validity of some handedness questionnaire items. *Neuropsychologia*, *12*(1), 43-47. doi:10.1016/0028-3932(74)90025-6.

Rader, C. M., & Tellegen, A. (1987). An investigation of synesthesia. Journal of Personality and Social Psychology, 52, 981-987.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia - A window into perception, thought and language. *Journal of Consciousness Studies*, 8(12), 3-34.

Ramachandran, V. S., & Hubbard, E. M. (2003). Hearing colors, tasting shapes. *Scientific American*, *288*(5), 52-59. doi:10.1038/scientificamerican0503-52.

Ramachandran, V. S., & Marcus, Z. (2017). Synesthesia and the McCollough Effect. *i-Perception*, *8*(3), 204166951771171. doi:10.1177/2041669517711718.

Ramos-Estebanez, C., Merabet, L. B., Machii, K., Fregni, F., Thut, G., Wagner, T. A., … Pascual-Leone, A. (2007). Visual phosphene perception modulated by subthreshold crossmodal sensory stimulation. *Journal of Neuroscience, 27*(15), 4178-4181. doi:10.1523/jneurosci.5468-06.2007.

Ratcliffe, V. F., Taylor, & A. M., Reby, D. (2016). Cross-Modal Correspondences in Non-human Mammal Communication. *Multisensory Research*, *29*(1-3), 49-91. doi:10.1163/22134808-00002509.

Ray, P. G., Meador, K. J., Epstein, C. M., Loring, D. W., & Day, L. J. (1998). Magnetic stimulation of visual cortex: Factors influencing the perception of phosphenes. *Journal of Clinical Neurophysiology, 15*(4), 351-357. doi:10.1097/00004691-199807000-00007.

Reetz, H., & Jongman, A. (2009). Phonetics: Transcription, production, acoustics, and perception. Chichester, U.K.: Blackwell.

Regenbogen, C., Seubert, J., Johannson, E. Finkelmeyer, A., Andersson, P. & Lundström, J.N. (2018). The intraparietal sulcus governs multisensory integration of audiovisual information based on task difficulty. *Human Brain Mapping*, 39:1313-1326.

Revill, K. P., Namy, L. L., Defife, L. C., & Nygaard, L. C. (2014). Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI. *Brain and Language*, 128(1), 18-24. doi: 10.1016/j.bandl.2013.11.002.

Revill, K. P., Namy, L. L., & Nygaard, L. C. (2018). Eye movements reveal persistent sensitivity to sound symbolism during word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(5), 680-698. doi:10.1037/xlm0000476.

Rich, A. N., & Mattingley, J. B. (2002). Anomalous perception in synaesthesia: A cognitive neuroscience perspective. *Nature Reviews Neuroscience, 3*(1), 43-52. doi:10.1038/nrn702.

Rich, A., Bradshaw, J., & Mattingley, J. (2005). A systematic, large-scale study of synaesthesia: Implications for the role of early experience in lexical-colour associations. *Cognition*, *98*(1), 53-84. doi:10.1016/j.cognition.2004.11.003.

Ricotta, C., & Pavoine, S. (2015). Measuring similarity among plots including similarity among species: an extension of traditional approaches. *Journal of Vegetation Science*, *26*(6), 1061-1067. doi:10.1111/jvs.12329.

Ro, T., Wallace, R., Hagedorn, J., Farné, A., & Pienkos, E. (2004). Visual enhancing of tactile perception in the posterior parietal cortex. *Journal of Cognitive Neuroscience, 16*(1), 24-30. doi:10.1162/089892904322755520.

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology*, *50*(1-2), 19-26. doi:10.1016/s0167-8760(03)00121-1.

Roffler, S.K. and Butler, R.A. (1968). Factors that influence the localization of sound in the vertical plane. *Journal of the Acoustical Society of America*, 43:1255-1259.

Rolls, E. T., Joliot, M., & Tzourio-Mazoyer, N. (2015). Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage*, *122*, 1-5. doi:10.1016/j.neuroimage.2015.07.075.

Romei, V., Murray, M. M., Cappe, C., & Thut, G. (2009). Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. *Current Biology*, *19*(21), 1799-1805. doi:10.1016/j.cub.2009.09.027.

Romei, V., Murray, M. M., Cappe, C., & Thut, G. (2013). The contributions of sensory dominance and attentional bias to cross-modal enhancement of visual cortex excitability. *Journal of Cognitive Neuroscience*, *25*(7), 1122-1135. doi:10.1162/jocn_a_00367.

Root, N. B., Rouw, R., Asano, M., Kim, C., Melero, H., Yokosawa, K., & Ramachandran, V. S. (2018). Why is the synesthete's “A” red? Using a five-language dataset to disentangle the effects of shape, sound, semantics, and ordinality on inducer–concurrent relationships in grapheme-color synesthesia. *Cortex*, *99*, 375-389. doi:10.1016/j.cortex.2017.12.003.

Rothen, N., Scott, R. B., Mealor, A. D., Coolbear, D. J., Burckhardt, V., & Ward, J. (2013). Synesthetic experiences enhance unconscious learning. *Cognitive Neuroscience*, *4*(3-4), 231-238. doi:10.1080/17588928.2013.847077.

Sachs (1812) Historiae Naturalis Duorum Leucaetiopum: Auctoris Ipsius et Sororis Eius. Erlangen.

Saffran, J. R., & Estes, K. G. (2006). Mapping sound to meaning: Connections between learning about sounds and learning about words. *Advances in Child Development and Behavior, 34*, 1–38.

Sagiv, N., Ilbeigi, I. & Ben-Tal, O. (2011). Reflections on synaesthesia, perception, and cognition. *Intellectica*, 55, 81-94.

Sandhu, R., & Dyson, B. J. (2012). Re-evaluating visual and auditory dominance through modality switching costs and congruency analyses. *Acta Psychologica*, *140*(2), 111-118. doi:10.1016/j.actpsy.2012.04.003.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, 12, 225–239.

Sathian, K., Simon, T.J., Peterson, S., Patel, G.A., Hoffman, J.M. & Grafton, S.T. (1999). Neural evidence linking visual object enumeration and attention. *Journal of Cognitive Neuroscience*, 11, 36-51.

Sathian, K., Zangaladze, A., Hoffman, J. M., & Grafton, S. T. (1997). Feeling with the mind's eye. *NeuroReport*, *8*(18), 3877-3881. doi:10.1097/00001756-199712220-00008.

Schmidtke, D.S., Conrad, M. & Jacobs, A.M. (2014). Phonological iconicity. *Frontiers in Psychology*, 5:80, doi: 10.3389/fpsyg.2014.00080.

Schneider,W., Eschman, A.,& Zuccolotto, A. (2002). *E-prime user's guide*. Pittsburgh: Psychology Software Tools Inc.

Schroeder, C. E., Smiley, J., Fu, K. G., McGinnis, T., O'Connell, M. N., & Hackett, T. A. (2003). Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *International Journal of Psychophysiology*, *50*(1-2), 5-17. doi:10.1016/s0167-8760(03)00120-x.

Segal, G. (1997). Synaesthesia: Implications for the modularity of mind. In S. Baron-Cohen & J.E. Harrison (Eds.), *Synaesthesia: Classic and contemporary reading* (pp. 211-224).Malden: Blackwell.

Seo, H. S., Arshamian, A., Schemmer, K., Scheer, I., Sander, T., Ritter, G., & Hummel, T. (2010). Crossmodal integration between odors and abstract symbols. *Neuroscience Letters*, *478*(3), 175–178. doi:10.1016/j.neulet.2010.05.011.

Sereno, J. A., & Jongman, A. (1990). Phonological and form class relations in the lexicon. *Journal of Psycholinguistic Research, 19*(6), 387–404. doi:10.1007/BF01068886.

Serino, A., Canzoneri, E., & Avenanti, A. (2011). Fronto-parietal areas necessary for a multisensory representation of peripersonal space in humans: An rTMS study. *Journal of Cognitive Neuroscience*, *23*(10), 2956-2967. doi:10.1162/jocn_a_00006.

Sestieri, C., Di Matteo, R., Ferretti, A., Del Gratta, C., Caulo, M., Tartaro, A., Belardinelli, M.O. & Romani, G.L. (2006). “What” versus “where” in the audiovisual domain: an fMRI study. *NeuroImage*, 33:672-680.

Šetic, M., & Domijan, D. (2007). The influence of vertical spatial orientation on property verification. *Language and Cognitive Processes*, *22*(2), 297–312. doi:10.1080/01690960600732430.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*(6814), 788-788. doi:10.1038/35048669.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants’ sensitivity to perceptual cues to lexical and grammatical words. *Cognition, 72*(2), B11–B21. doi:10.1016/S0010-0277(99)00047-5.

Shinohara, K., & Kawahara, S. (2010). A cross-linguistic study of sound symbolism: The images of size. *Annual Meeting of the Berkeley Linguistics Society, 36*(1), 396. doi:10.3765/bls.v36i1.3926.

Shor, R.E. (1975). An auditory analog of the Stroop test. *Journal of General Psychology*, 93:281-288.

Shrum, L., Lowrey, T., Luna, D., Lerman, D., & Liu, M. (2012). Sound symbolism effects across languages: Implications for global brand names. *International Journal of Research in Marketing*, 29(3), 275-279. doi: 10.1016/j.ijresmar.2012.03.002

Sidhu, D. M., & Pexman, P. M. (2017). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-017-1361-1.

Simmons, W. K., Hamann, S. B., Harenski, C. L., Hu, X. P., & Barsalou, L. W. (2008). FMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology-Paris*, 102(1-3), 106-119. doi: 10.1016/j.jphysparis.2008.03.014

Simner, J. (2007). Beyond perception: synaesthesia as a psycholinguistic phenomenon. *Trends in Cognitive Sciences*, *11*(1), 23-29. doi:10.1016/j.tics.2006.10.010

Simner, J., & Bain, A. E. (2017). Do children with grapheme-colour synaesthesia show cognitive benefits? *British Journal of Psychology*, *109*(1), 118-136. doi:10.1111/bjop.12248

Simner, J., & Carmichael, D. A. (2015). Is synaesthesia a dominantly female trait? *Cognitive Neuroscience*, *6*(2-3), 68–76. doi: 10.1080/17588928.2015.1019441

Simner, J., Ipser, A., Smees, R., & Alvarez, J. (2017). Does synaesthesia age? Changes in the quality and consistency of synaesthetic associations. *Neuropsychologia*, *106*, 407-416. doi:10.1016/j.neuropsychologia.2017.09.013

Sinke, C., Neufeld, J., Emrich, H., Dillo, W., Bleich, S., Zedler, M., & Szycik, G. (2012). Inside a synesthete's head: A functional connectivity analysis with grapheme-color synesthetes. *Neuropsychologia*, *50*(14), 3363-3369. doi:10.1016/j.neuropsychologia.2012.09.015.

Simner, J., Harrold, J., Creed, H., Monro, L., & Foulkes, L. (2009). Early detection of markers for synaesthesia in childhood populations. *Brain*, *132*(1), 57-64. doi:10.1093/brain/awn292.

Simner, J., & Holenstein, E. (2007). Ordinal linguistic personification as a variant of synesthesia. *Journal of Cognitive Neuroscience*, *19*(4), 694-703. doi:10.1162/jocn.2007.19.4.694.

Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L., & Oakley, D. A. (2005). Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations. *Cognitive Neuropsychology*, *22*(8), 1069–1085. doi:10.1080/02643290500200122.

Simner, J., & Ludwig, V. U. (2012). The color of touch: A case of tactile–visual synaesthesia. *Neurocase*, *18*(2), 167-180. doi:10.1080/13554794.2011.568503.

Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., … Ward, J. (2006). Synaesthesia: The prevalence of atypical cross-modal experiences. *Perception*, *35*(8), 1024-1033. doi:10.1068/p5469.

Singh, L. (2015). Speech signal analysis using FFT and LPC. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 4*(4), 1658-1660.

Smiley, J. F., Hackett, T. A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D. C., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *The Journal of Comparative Neurology*, *502*(6), 894-923. doi:10.1002/cne.21325.

Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as process: Contextually cued attention and early word learning. *Cognitive Science*, *34*, 1287–1314.

Smith, L.B. & Sera M.D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, 24:99-142.

Spector, F., & Maurer, D. (2008). The colour of Os: Naturally biased associations between shape and colour. *Perception*, *37*(6), 841–847. doi:10.1068/p5830.

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

Spence, C. (2012). Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism. *Journal of Consumer Psychology,* 22(1), 37-54.

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences*, 1296:31-49.

Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition*, *22*(1), 245-260. doi:10.1016/j.concog.2012.12.006

Spence, C., & Deroy, O. (2013). On why music changes what (we think) we taste. *I-Perception,* 4(2), 137-140. doi: 10.1068/i0577ic.

Spence, C., Wan, X., Woods, A., Velasco, C., Deng, J., Youssef, J., & Deroy, O. (2015). On tasty colours and colourful tastes? Assessing, explaining, and utilizing crossmodal correspondences between colours and basic tastes. *Flavour*, *4*(1). doi:10.1186/s13411-015-0033-1.

Sperdin, H. F., Cappe, C., & Murray, M. M. (2010). Auditory–somatosensory multisensory interactions in humans: Dissociating detection and spatial discrimination. *Neuropsychologia*, *48*(13), 3696-3705. doi:10.1016/j.neuropsychologia.2010.09.001.

Stanovich, K. E. (1992). Speculations on the causes and consequences of individual differences in early reading acquisition. In P.B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 307–342). Hillsdale, NJ: Lawrence Erlbaum.

Staub, A., Grant, M., Clifton, C., & Rayner, K. (2009). Phonological typicality does not influence fixation durations in normal reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 806–814.

Stein, B. E., Meredith, M. A., Huneycutt, W. S., & McDade, L. (1989). Behavioral indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli. *Journal of Cognitive Neuroscience*, *1*(1), 12-24. doi:10.1162/jocn.1989.1.1.12.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press (MA).

Stevenson, R.A., Altieri, N.A., Kim, S., Pisoni, D.B. & James, T.W. (2010). Neural processing of asynchronous audiovisual speech perception. *NeuroImage*, 49:3308-3318.

Stevenson, R. A., Baum, S. H., Krueger, J., Newhouse, P. A., & Wallace, M. T. (2018). Links between temporal acuity and multisensory integration across life span. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 106-116. doi:10.1037/xhp0000424.

Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., … Wallace, M. T. (2014). Identifying and quantifying multisensory integration: A tutorial review. *Brain Topography*, *27*(6), 707-730. doi:10.1007/s10548-014-0365-7.

Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research*, *227*(2), 249-261. doi:10.1007/s00221-013-3507-3.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643-662. doi:10.1037/h0054651.

Styles, S.J. & Gawne, L. (2017). When does maluma/takete fail? Two key failures and a meta-analysis suggest that phonology and phonotactics matter. *i-Perception,* doi:10.1177/2041669517724807

Sučević, J., Savić, A.M., Popović, M.B., Styles, S.J. & Ković, V. (2015). Balloons and bavoons versus spikes and shikes: ERPs reveal shared neural processes for shape-sound-meaning congruence in words, and shape-sound congruence in pseudowords. *Brain & Language*, 145/146:11-22.

Talairach, J. & Tournoux, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain*. Thieme Medical Publishers; New York.

Teixeira, J. P., & Fernandes, P. O. (2014). Jitter, shimmer and HNR classification within gender, tones and vowels in healthy voices. *Procedia Technology, 16*, 1228-1237. doi:10.1016/j.protcy.2014.10.138.

Thompson, P.D. & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *Quarterly Journal of Experimental Psychology*, 64:2392-2404.

Tilot, A. K., Kucera, K. S., Vino, A., Asher, J. E., Baron-Cohen, S., & Fisher, S. E. (2018). Rare variants in axonogenesis genes connect three families with sound–color synesthesia. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(12), 3168–3173. doi: 10.1073/pnas.1715492115.

Todd, M. T., Nystrom, L. E., & Cohen, J. D. (2013). Confounds in multivariate pattern analysis: Theory and rule representation case study. *NeuroImage*, *77*, 157-165. doi:10.1016/j.neuroimage.2013.03.039.

Tolar, T. D., Lederberg, A. R., Gokhale, S., & Tomasello, M. (2008). The development of the ability to recognize the meaning of iconic signs. *Journal of Deaf Studies and Deaf Education, 13,* 225–240.

Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual Review of Psychology, 63*(1), 483-509. doi:10.1146/annurev-psych-120710-100412.

Topolinski, S., Maschmann, I. T., Pecher, D., & Winkielman, P. (2014). Oral approach-avoidance: Affective consequences of muscular articulation dynamics. *Journal of Personality and Social Psychology*, 106(6), 885-896.

Tyler, L.K., Bright, P., Fletcher, P. & Stamatakis, E.A. (2004). Neural processing of nouns and verbs: the role of inflectional morphology. *Neuropsychologia*, 42:512-523.

Tyll, S., Bonath, B., Schoenfeld, M. A., Heinze, H., Ohl, F. W., & Noesselt, T. (2013). Neural basis of multisensory looming signals. *NeuroImage*, *65*, 13-22. doi:10.1016/j.neuroimage.2012.09.056.

Tzeng, C. Y., Duan, J., Namy, L. L., & Nygaard, L. C. (2018). Prosody in speech as a source of referential information. *Language, Cognition and Neuroscience*, *33*(4), 512-526. doi:10.1080/23273798.2017.1391400.

Tzeng, C. Y., Nygaard, L. C., & Namy, L. L. (2016). The Specificity of Sound Symbolic Correspondences in Spoken Language. *Cognitive Science*, *41*(8), 2191-2220. doi:10.1111/cogs.12474.

Tzeng, C. Y., Nygaard, L. C., & Namy, L. L. (2017). Developmental change in children's sensitivity to sound symbolism. *Journal of Experimental Child Psychology,160*, 107-118. doi:10.1016/j.jecp.2017.03.004.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject Brain. *NeuroImage, 15*(1), 273-289. doi:10.1006/nimg.2001.0978.

Vainio, L., Tiainen, M., Tiippana, K., Rantala, A., & Vainio, M. (2016). Sharp and round shapes of seen objects have distinct influences on vowel and consonant articulation. *Psychological Research, 81*(4), 827-839. doi:10.1007/s00426-016-0778-x

van Leeuwen, T. M., Trautmann-Lengsfeld, S. A., Wallace, M. T., Engel, A. K., & Murray, M. M. (2016). Bridging the gap: Synaesthesia and multisensory processes. *Neuropsychologia*, *88*, 1-4. doi:10.1016/j.neuropsychologia.2016.06.007.

van Atteveldt, N.M., Formisano, E., Blomert, L. & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex*, 17:962-974.

van Atteveldt, N., Murray, M., Thut, G., & Schroeder, C. (2014). Multisensory integration: Flexible use of general operations. *Neuron, 81*(6), 1240-1253. doi:10.1016/j.neuron.2014.02.044.

van Atteveldt, N. M., Peterson, B. S., & Schroeder, C. E. (2013). Contextual control of audiovisual integration in low-level sensory cortices. *Human Brain Mapping*, *35*(5), 2394-2411. doi:10.1002/hbm.22336.

Van der Burg, E., Alais, D., & Cass, J. (2013). Rapid recalibration to audiovisual asynchrony. *Journal of Neuroscience, 33*(37), 14633-14637. doi:10.1523/jneurosci.1182-13.2013.

Venezia, J. H., Vaden, K. I., Rong, F., Maddox, D., Saberi, K., & Hickok, G. (2017). Auditory, visual and audiovisual speech processing streams in superior temporal sulcus. *Frontiers in Human Neuroscience*, *11*. doi:10.3389/fnhum.2017.00174.

Vetter, P., Smith, F., & Muckli, L. (2014). Decoding sound and imagery content in early visual cortex. *Current Biology*, *24*(11), 1256-1262. doi:10.1016/j.cub.2014.04.020.

Vroomen, J., & Baart, M. (2011). Phonetic recalibration in audiovisual speech. *Frontiers in Neuroscience*, 363-380. doi:10.1201/b11092-24.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2009). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science, 21*(1), 21–25. doi:10.1177/0956797609354734.

Walker, P., & Smith, S. (1985). Stroop interference based on the multimodal correlates of haptic size and auditory pitch. *Perception*, *14*(6), 729–736. doi:10.1068/p140729.

Walker, L., Walker, P. and Francis, B. (2012). A common scheme for cross-sensory correspondences across stimulus domains. *Perception*, 41:1186-1192.

Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*, *13*, 286–350.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, 7:483-488.

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, *137*, 188-200. doi:10.1016/j.neuroimage.2015.12.012.

Wan, X., Woods, A. T., van den Bosch, J. J., McKenzie, K. J., Velasco, C., & Spence, C. (2014). Crosscultural differences in crossmodal correspondences between basic tastes and visual features. *Frontiers in Psychology*, *5*, 1365. doi:10.3389/fpsyg.2014.01365.

Wang, Q. J., Wang, S., & Spence, C. (2016). “Turn up the taste”: Assessing the role of taste intensity and emotion in mediating crossmodal correspondences between basic tastes and pitch. *Chemical Senses*, *41*(4), 345-356. doi:10.1093/chemse/bjw007.

Ward, J., Hoadley, C., Hughes, J. E. A., Smith, P., Allison, C., Baron-Cohen, S., & Simner, J. (2017). Atypical sensory sensitivity as a shared feature between synaesthesia and autism. *Scientific Reports*, *7*, 41155. doi: 10.1038/srep41155.

Ward, J., & Simner, J. (2003). Lexical-gustatory synaesthesia: linguistic and conceptual factors. *Cognition*, *89*(3), 237-261. doi:10.1016/s0010-0277(03)00122-7.

Ward, J., & Simner, J. (2003). Lexical-gustatory synaesthesia: linguistic and conceptual factors. *Cognition*, *89*(3), 237-261. doi:10.1016/s0010-0277(03)00122-7.

Ward, J., Simner, J., & Auyeung, V. (2005). A comparison of lexical-gustatory and grapheme-colour synaesthesia. *Cognitive Neuropsychology*, *22*(1), 28-41. doi:10.1080/02643290442000022.

Watkins, S., Shams, L., Josephs, O., & Rees, G. (2007). Activity in human V1 follows multisensory perception. *NeuroImage*, *37*(2), 572-578. doi:10.1016/j.neuroimage.2007.05.027.

Watson, M. R., Chromý, J., Crawford, L., Eagleman, D. M., Enns, J. T., & Akins, K. A. (2017). The prevalence of synaesthesia depends on early language learning. *Consciousness and Cognition*, *48*, 212-231. doi:10.1016/j.concog.2016.12.004.

Weiss, J. H. (1966). A study of the ability of English speakers to guess the meanings of non-antonym foreign words. *Journal of General Psychology, 74*(1), 97–106. doi:10.1080/00221309.1966.9710313.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*(3), 638-667. doi:10.1037//0033-2909.88.3.638.

Werner, S., & Noppeney, U. (2010a). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *Journal of Neuroscience, 30*(7), 2662-2675. doi:10.1523/jneurosci.5091-09.2010.

Werner, S. & Noppeney, U. (2010b). Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cerebral Cortex*, 20:1829-1842.

Westbury, C. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language*, *93*(1), 10–19. doi:10.1016/j.bandl.2004.07.006.

Westbury, C., Hollis, G., Sidhu, D. M., & Pexman, P. M. (2018). Weighing up the evidence for sound symbolism: Distributional properties predict cue strength. *Journal of Memory and Language*, *99*, 122-150. doi:10.1016/j.jml.2017.09.006

White, T., O'Leary, D., Magnotta, V., Arndt, S., Flaum, M. & Andreasen, N.C. (2001). Anatomic and functional variability: The effects of filter size in group fMRI data analysis. *NeuroImage*, 13:577-588.

Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology, 4*. doi:10.3389/fpsyg.2013.00058.

Wilson, L.B., Tregellas, J.R., Slason, E., Pasko, B.E. & Rojas, D.C. (2011). Implicit phonological priming during visual word recognition. *NeuroImage*, 55:724-731.

Witthoft, N., & Winawer, J. (2006). Synesthetic colors determined by having colored refrigerator magnets in childhood. *Cortex*, *42*(2), 175-183. doi:10.1016/s0010-9452(08)70342-3.

Woo, C.-W., Krishnan, A. & Wager, T.D. (2014). Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. *NeuroImage*, 91:412-419.

Wordsworth, W. (1910). *The Complete Poetical Works of William Wordsworth: 1806-1815*.

Wurm, L. H. (2007). Danger and usefulness: An alternative framework for understanding rapid evaluation effects in perception. *Psychonomic Bulletin and Review, 14*(6), 1218–1225. doi:10.3758/BF03193116.

Xie, Y., Xu, Y., Bian, C. & Li, M. (2017). Semantic congruent audiovisual integration during the encoding stage of working memory: an ERP and sLORETA study. *Scientific Reports*, 7: 5112, doi:10.1038/s41598-017-05471-1.

Yau, J. M., DeAngelis, G. C., & Angelaki, D. E. (2015). Dissecting neural circuits for multisensory integration and crossmodal processing. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1677), 20140203. doi:10.1098/rstb.2014.0203.

Yokoyama, T., Noguchi, Y., Koga, H., Tachibana, R., Saiki, J., Kakigi, R., & Kita, S. (2014). Multiple neural mechanisms for coloring words in synesthesia. *NeuroImage*, *94*, 360-371. doi:10.1016/j.neuroimage.2014.01.039.

Yoshida, H. (2012). A cross-linguistic study of sound symbolism in children’s verb learning. *Journal of Cognition and Development, 13*(2), 232–265. doi:10.1080/15248372.2011.573515.

Yoshida, H., & Smith, L. B. (2003). Sound symbolism and early word learning in two languages. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Cognitive Science Society Conference*, 1287–1292, Mahwah, NJ: Lawrence Erlbaum.

Zangaladze, A., Epstein, C. M., Grafton, S. T., & Sathian, K. (1999). Involvement of visual cortex in tactile discrimination of orientation. *Nature*, *401*(6753), 587-590. doi:10.1038/44139.

Zangenehpour, S., & Zatorre, R. J. (2010). Crossmodal recruitment of primary visual cortex following brief exposure to bimodal audiovisual stimuli. *Neuropsychologia*, *48*(2), 591-600. doi:10.1016/j.neuropsychologia.2009.10.022.

Žarić, G., Fraga González, G., Tijms, J., Van der Molen, M. W., Blomert, L., & Bonte, M. (2015). Crossmodal deficit in dyslexic children: Practice affects the neural timing of letter-speech sound integration. *Frontiers in Human Neuroscience*, *9*. doi:10.3389/fnhum.2015.00369.

Zhang, Y., Peng, B., Wang, S., Liang, Y., Yang, J., So, K., & Yuan, T. (2016). Image processing methods to elucidate spatial characteristics of retinal microglia after optic nerve transection. *Scientific Reports, 6*(1). doi:10.1038/srep21816.

Zhang, Z., Sun, Y., Humphreys, G.W. & Song, Y. (2017). Different activity patterns for action and language within their shared neural areas: an fMRI study on action observation and language phonology. *Neuropsychologia*, 99:112-120.

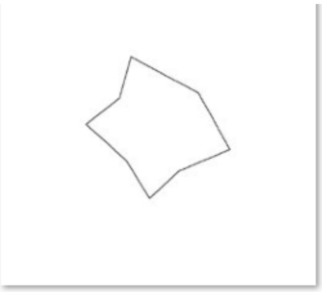
Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 35–62). New York: Academic Press.

Appendix 1

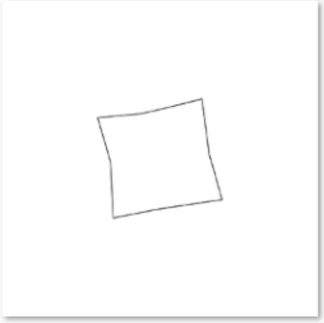
Visual Shape Stimuli for Experiments

<u>Pointed Shapes:</u>

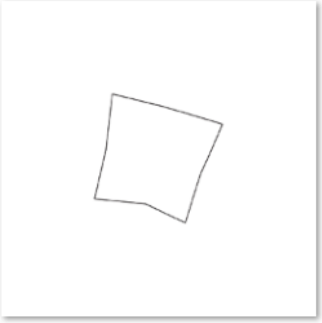

A1


A2


A3


A4

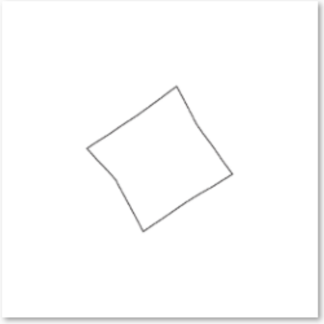

A5


A6


A7

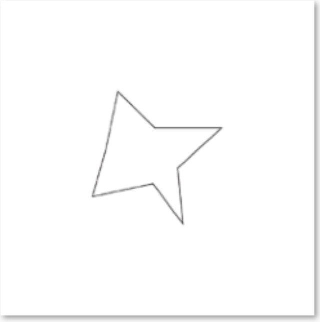

A8


A9

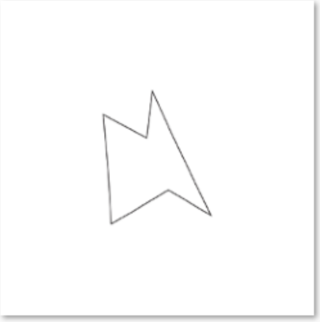

A10


A11

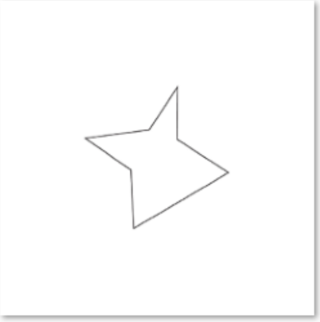

A12



A13

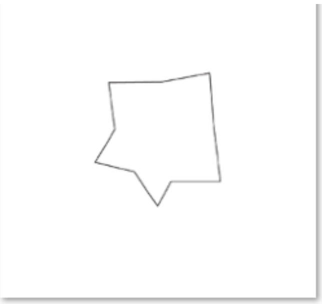


A14



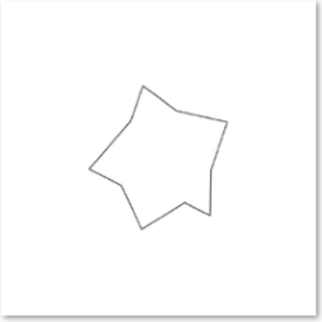
A15

C1



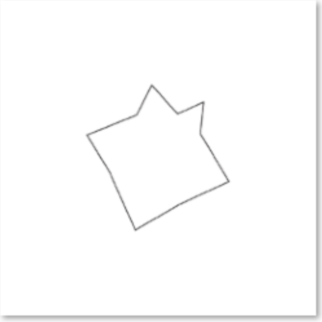
C2

C3



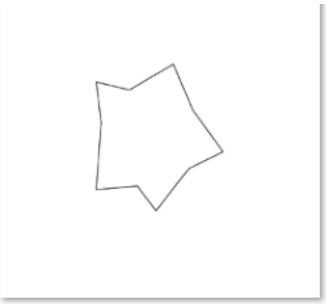
C4

C5

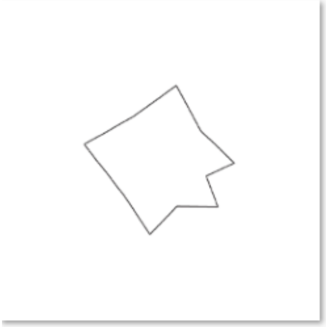


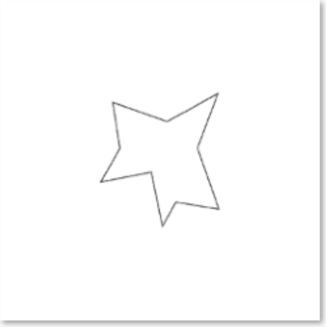
C6


C7

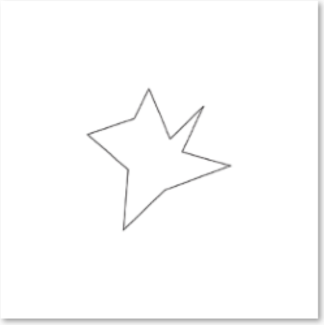

C8

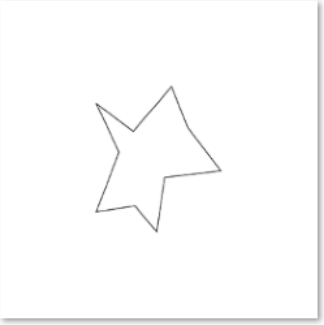

C9


C10


C11


C12


C13

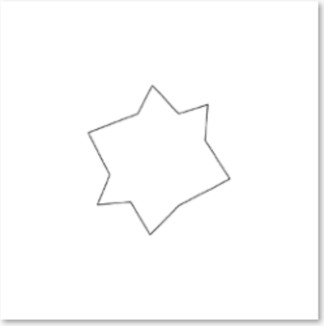

C14


C15

E1

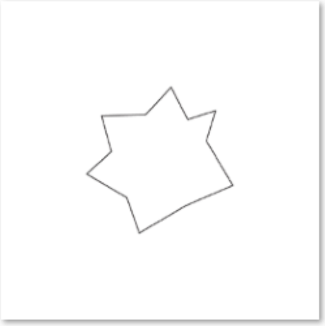
E2

E3



E4

E5

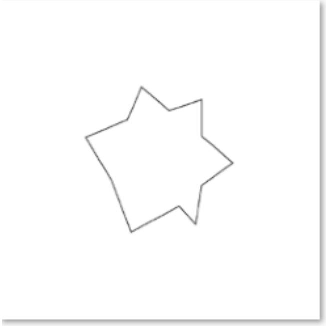


E6

E7

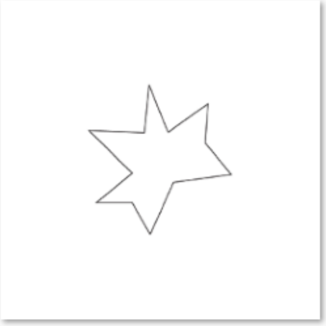
E8

E9



E10

E11



E12

E13

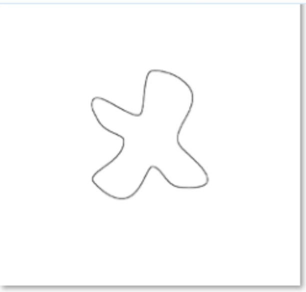
E14

E15

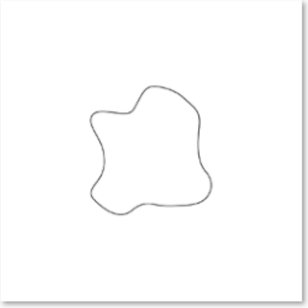
Rounded Shapes:

B1

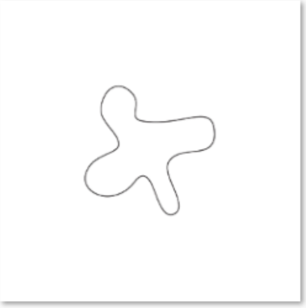
B2



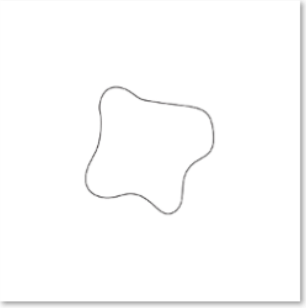
B3



B4



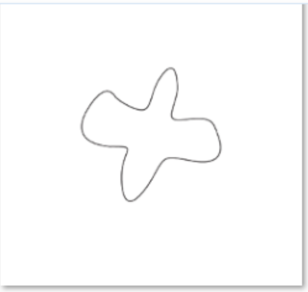
B5

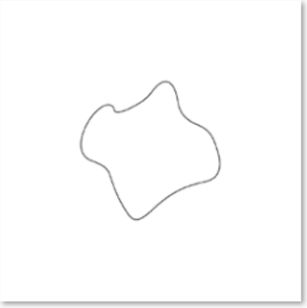


B6

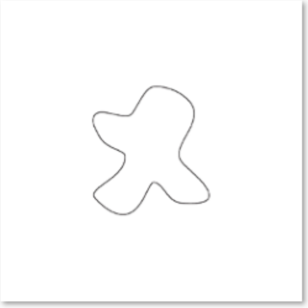

B7

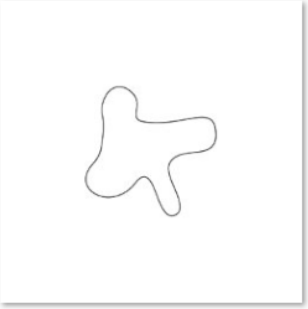

B8

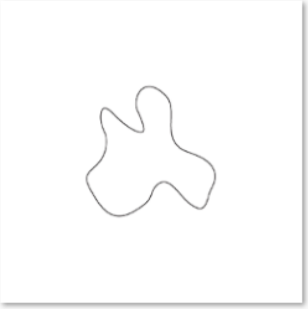

B9

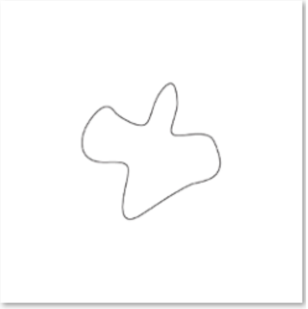

B10


B11

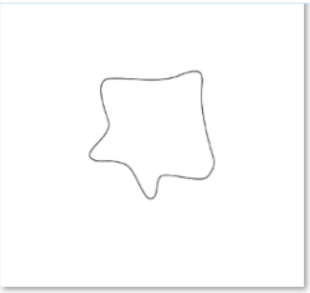

B12

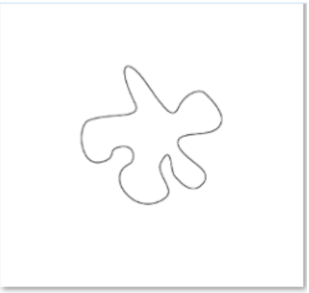

B13

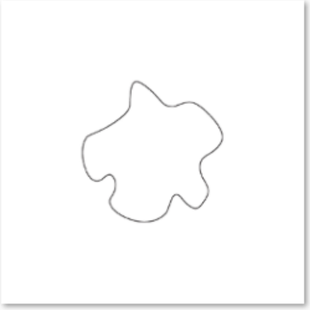

B14

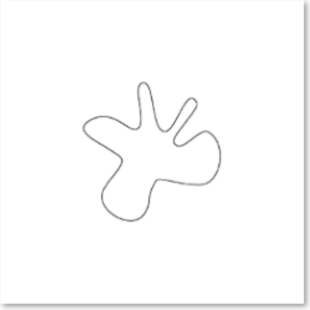

B15

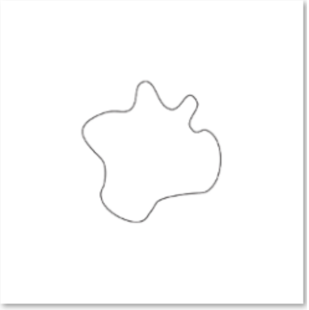

D1


D2

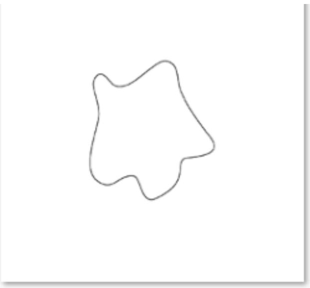

D3

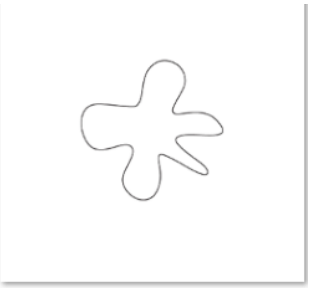

D4

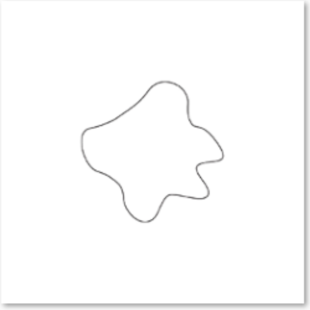

D5

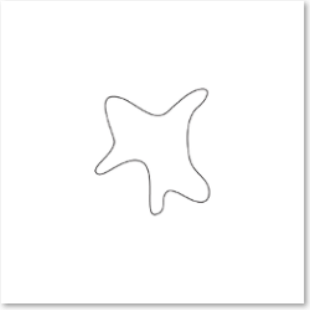

D6

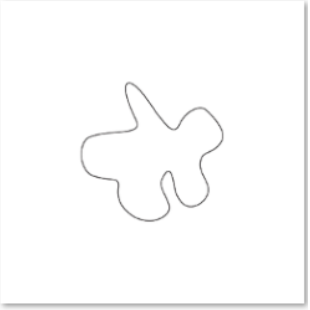

D7

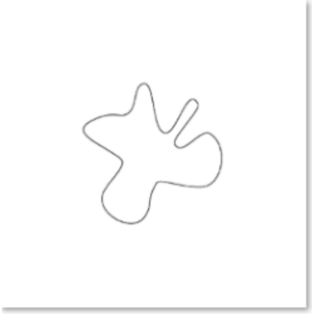

D8

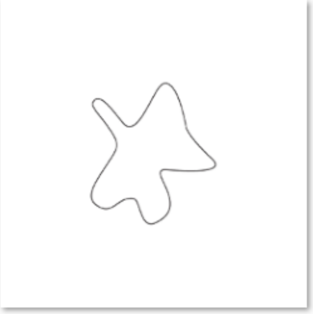

D9

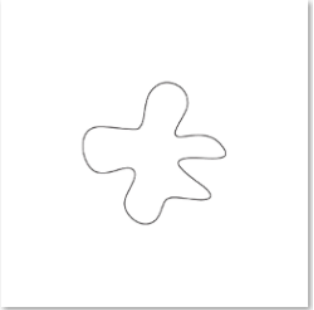

D10


D11


D12

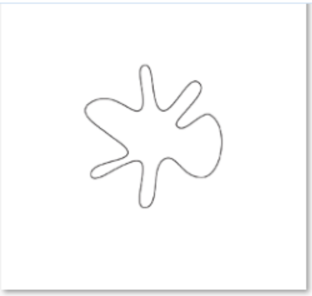

D13


D14

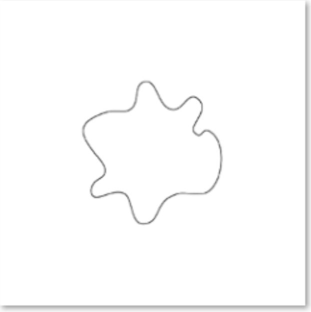

D15

F1

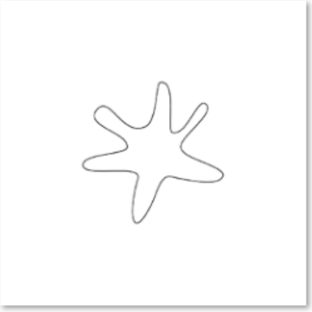
F2



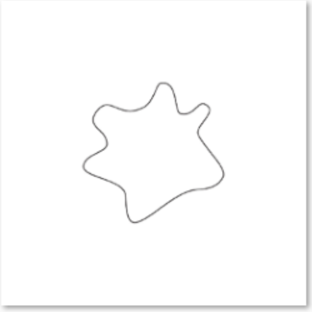
F3



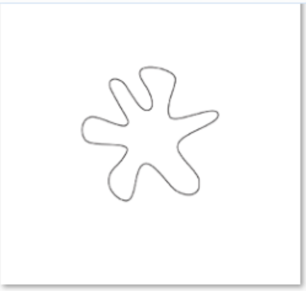
F4



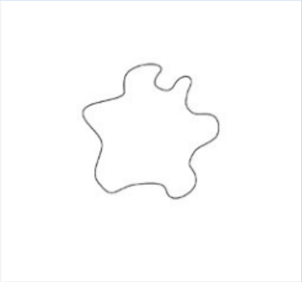
F5



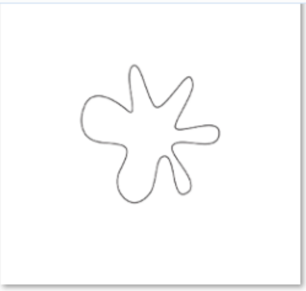
F6



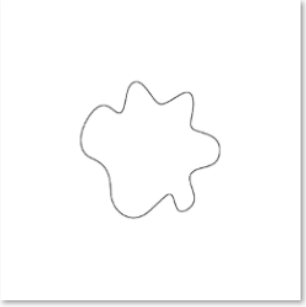
F7



F8

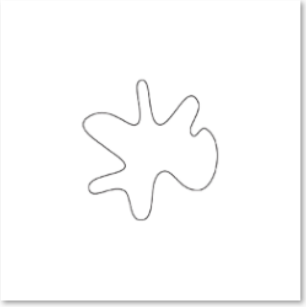


F9

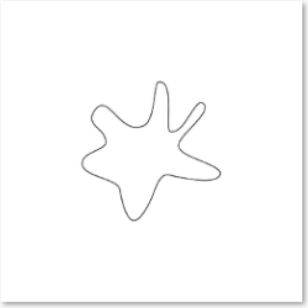


F10

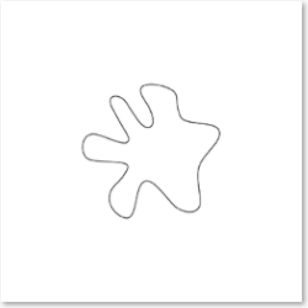
F11



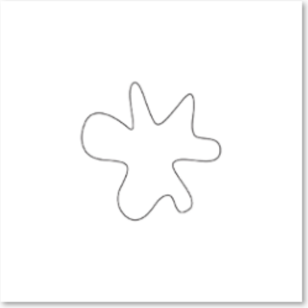
F12



F13



F14



F15

Appendix 2

Auditory Pseudoword Stimuli for Experiments

Sonorant Consonant 1	**Voiced Vowel 1**	**Unrounded Consonant 2**	**Front Vowel 2**	**Sonorant Consonant 1**	**Voiced Vowel 1**	**Rounded Consonant 2**	**Back Vowel 2**
l	e	l	i	l	o	l	u
l	ɛ	l	e	l	u	l	o
l	ɛ	l	i	l	ʊ	l	o
l	i	l	e	l	ʊ	l	u
l	I	l	e	l	o	m	o
l	e	m	e	l	o	m	u
l	e	m	i	l	u	m	o
l	ɛ	m	e	l	u	m	u
l	ɛ	m	i	l	ʊ	m	o
l	i	m	e	l	ʊ	m	u
l	I	m	e	l	o	n	o
l	i	m	i	l	o	n	u
l	I	m	i	l	u	n	o
l	e	n	e	l	u	n	u
l	e	n	i	l	ʊ	n	o
l	ɛ	n	e	l	ʊ	n	u
l	i	n	e	m	o	l	o
l	I	n	e	m	o	l	u
l	i	n	i	m	u	l	o
l	I	n	i	m	u	l	u
m	e	l	i	m	ʊ	l	o
m	ɛ	l	e	m	ʊ	l	u
m	ɛ	l	i	m	o	m	u
m	i	l	e	m	u	m	o
m	I	l	e	m	ʊ	m	o
m	i	l	i	m	ʊ	m	u
m	e	m	i	m	o	n	o
m	ɛ	m	e	m	o	n	u
m	ɛ	m	i	m	u	n	o
m	i	m	e	m	u	n	u
m	I	m	e	m	ʊ	n	o
m	I	m	i	m	ʊ	n	u
m	e	n	e	n	o	l	o
m	e	n	i	n	o	l	u
m	ɛ	n	e	n	u	l	o
m	i	n	e	n	u	l	u

Sonorant Consonant 1	Voiced Vowel 1	Unrounded Consonant 2	Front Vowel 2
m	I	n	e
n	e	l	e
n	e	l	i
n	ɛ	l	e
n	i	l	e
n	I	l	e
n	i	l	i
n	I	l	i
n	e	m	e
n	e	m	i
n	ɛ	m	e
n	ɛ	m	i
n	i	m	e
n	I	m	e
n	i	m	i
n	I	m	i
n	e	n	i
n	ɛ	n	e
n	ɛ	n	i
n	i	n	e
n	I	n	e

Sonorant Consonant 1	Voiced Vowel 1	Rounded Consonant 2	Back Vowel 2
n	ʊ	l	o
n	ʊ	l	u
n	o	m	o
n	o	m	u
n	u	m	o
n	u	m	u
n	ʊ	m	o
n	ʊ	m	u
n	o	n	u
n	u	n	o
n	ʊ	n	o
n	ʊ	n	u

Stops Consonant 1	Voiced Vowel 1	Unrounded Consonant 2	Front Vowel 2
b	ɛ	b	e
b	ɛ	b	i
b	i	b	e
b	I	b	e
b	I	b	i
b	e	d	e
b	e	d	i
b	ɛ	d	e
b	ɛ	d	i
b	i	d	e
b	e	g	e
b	e	g	i
b	ɛ	g	e
b	ɛ	g	i
b	i	g	e
b	I	g	e
b	i	g	i

Stops Consonant 1	Voiced Vowel 1	Rounded Consonant 2	Back Vowel 2
b	o	b	u
b	u	b	o
b	ʊ	b	o
b	ʊ	b	u
b	o	d	o
b	o	d	u
b	u	d	o
b	u	d	u
b	ʊ	d	o
b	ʊ	d	u
b	o	g	o
b	o	g	u
b	u	g	o
b	u	g	u
b	ʊ	g	o
b	ʊ	g	u
d	o	b	o

Stops	**Voiced**	**Unrounded**	**Front**
Consonant 1	**Vowel 1**	**Consonant 2**	**Vowel 2**
d	e	b	e
d	e	b	i
d	ε	b	e
d	i	b	e
d	I	b	e
d	i	b	i
d	I	b	i
d	e	d	i
d	ε	d	e
d	ε	d	i
d	i	d	e
d	I	d	e
d	e	g	e
d	e	g	i
d	ε	g	e
d	ε	g	i
d	i	g	e
d	I	g	e
d	i	g	i
d	I	g	i
g	e	b	e
g	e	b	i
g	ε	b	e
g	ε	b	i
g	i	b	e
g	I	b	e
g	i	b	i
g	I	b	i
g	e	d	e
g	e	d	i
g	ε	d	e
g	ε	d	i
g	i	d	e
g	I	d	e
g	i	d	i
g	e	g	i
g	ε	g	e
g	ε	g	i
g	i	g	e
g	I	g	e

Stops	**Voiced**	**Rounded**	**Back**
Consonant 1	**Vowel 1**	**Consonant 2**	**Vowel 2**
d	o	b	u
d	u	b	o
d	u	b	u
d	ʊ	b	o
d	ʊ	b	u
d	o	d	u
d	u	d	o
d	ʊ	d	o
d	ʊ	d	u
d	o	g	o
d	o	g	u
d	u	g	o
d	u	g	u
d	ʊ	g	o
d	ʊ	g	u
g	o	b	o
g	o	b	u
g	u	b	o
g	u	b	u
g	ʊ	b	o
g	ʊ	b	u
g	o	d	o
g	o	d	u
g	u	d	o
g	u	d	u
g	ʊ	d	o
g	ʊ	d	u
g	o	g	u
g	u	g	o
g	ʊ	g	o
g	ʊ	g	u

Stops	Voiced	Unrounded	Front
Consonant 1	**Vowel 1**	**Consonant 2**	**Vowel 2**
g	I	g	i

Stops	Unvoiced	Unrounded	Front	Stops	Unvoiced	Rounded	Back
Consonant 1	**Vowel 1**	**Consonant 2**	**Vowel 2**	**Consonant 1**	**Vowel 1**	**Consonant 2**	**Vowel 2**
k	e	k	i	k	o	k	u
k	ɛ	k	e	k	u	k	o
k	ɛ	k	i	k	ʊ	k	o
k	i	k	e	k	ʊ	k	u
k	I	k	e	k	o	p	o
k	I	k	i	k	o	p	u
k	e	p	e	k	u	p	o
k	e	p	i	k	u	p	u
k	ɛ	p	e	k	ʊ	p	o
k	ɛ	p	i	k	ʊ	p	u
k	i	p	e	k	o	t	o
k	I	p	e	k	o	t	u
k	i	p	i	k	u	t	o
k	I	p	i	k	u	t	u
k	e	t	e	k	ʊ	t	o
k	ɛ	t	e	k	ʊ	t	u
k	ɛ	t	i	p	o	k	o
k	i	t	e	p	o	k	u
k	I	t	e	p	u	k	o
k	i	t	i	p	u	k	u
p	e	k	e	p	ʊ	k	o
p	e	k	i	p	ʊ	k	u
p	ɛ	k	e	p	o	p	u
p	ɛ	k	i	p	u	p	o
p	i	k	e	p	ʊ	p	o
p	I	k	e	p	ʊ	p	u
p	i	k	i	p	o	t	o
p	e	p	i	p	o	t	u
p	i	p	e	p	u	t	o
p	I	p	e	p	u	t	u
p	e	t	e	p	ʊ	t	o
p	e	t	i	p	ʊ	t	u
p	ɛ	t	e	t	o	k	o
p	i	t	e	t	o	k	u
p	I	t	e	t	u	k	o
p	i	t	i	t	u	k	u

Stops Consonant 1	Unvoiced Vowel 1	Unrounded Consonant 2	Front Vowel 2
t	e	k	e
t	e	k	i
t	ɛ	k	e
t	i	k	e
t	I	k	e
t	I	k	i
t	e	p	e
t	e	p	i
t	ɛ	p	e
t	ɛ	p	i
t	i	p	e
t	I	p	e
t	I	p	i
t	e	t	i
t	ɛ	t	e
t	ɛ	t	i
t	i	t	e
t	I	t	e
k	e	k	i
k	ɛ	k	e
k	ɛ	k	i
k	i	k	e
k	I	k	e
k	I	k	i
k	e	p	e
k	e	p	i
k	ɛ	p	e
k	ɛ	p	i
k	i	p	e

Stops Consonant 1	Unvoiced Vowel 1	Rounded Consonant 2	Back Vowel 2
t	ʊ	k	o
t	ʊ	k	u
t	o	p	o
t	o	p	u
t	u	p	o
t	u	p	u
t	ʊ	p	o
t	ʊ	p	u
t	o	t	u
t	u	t	o
t	ʊ	t	o
t	ʊ	t	u

Affricate/ Fricatives Consonant 1	Voiced Vowel 1	Unrounded Consonant 2	Front Vowel 2
dʒ	e	dʒ	i
dʒ	ɛ	dʒ	e
dʒ	ɛ	dʒ	i
dʒ	i	dʒ	e
dʒ	I	dʒ	e
dʒ	I	dʒ	i
dʒ	e	v	e
dʒ	e	v	i

Affricate/ Fricatives Consonant 1	Voiced Vowel 1	Rounded Consonant 2	Back Vowel 2
dʒ	o	dʒ	u
dʒ	u	dʒ	o
dʒ	ʊ	dʒ	o
dʒ	ʊ	dʒ	u
dʒ	o	v	o
dʒ	o	v	u
dʒ	u	v	o
dʒ	u	v	u

Affricate/ Fricatives	Voiced	Unrounded	Front	Affricate/ Fricatives	Voiced	Rounded	Back
Consonant 1	Vowel 1	Consonant 2	Vowel 2	Consonant 1	Vowel 1	Consonant 2	Vowel 2
dʒ	ɛ	v	e	dʒ	ʊ	v	o
dʒ	ɛ	v	i	dʒ	ʊ	v	u
dʒ	i	v	e	dʒ	o	z	o
dʒ	I	v	e	dʒ	o	z	u
dʒ	i	v	i	dʒ	u	z	o
dʒ	I	v	i	dʒ	u	z	u
dʒ	e	z	e	dʒ	ʊ	z	o
dʒ	ɛ	z	e	dʒ	ʊ	z	u
dʒ	ɛ	z	i	v	o	dʒ	o
dʒ	i	z	e	v	o	dʒ	u
dʒ	I	z	e	v	u	dʒ	o
dʒ	I	z	i	v	u	dʒ	u
v	e	dʒ	e	v	ʊ	dʒ	o
v	e	dʒ	i	v	ʊ	dʒ	u
v	ɛ	dʒ	e	v	o	v	u
v	i	dʒ	e	v	u	v	o
v	I	dʒ	e	v	ʊ	v	o
v	i	dʒ	i	v	ʊ	v	u
v	I	dʒ	i	v	o	z	o
v	e	v	i	v	o	z	u
v	ɛ	v	e	v	u	z	o
v	ɛ	v	i	v	u	z	u
v	i	v	e	v	ʊ	z	o
v	I	v	e	v	ʊ	z	u
v	I	v	i	z	o	dʒ	o
v	e	z	e	z	o	dʒ	u
v	e	z	i	z	u	dʒ	o
v	ɛ	z	e	z	u	dʒ	u
v	ɛ	z	i	z	ʊ	dʒ	o
v	i	z	e	z	ʊ	dʒ	u
v	I	z	e	z	o	v	o
v	i	z	i	z	o	v	u
v	I	z	i	z	u	v	o
z	e	dʒ	e	z	u	v	u
z	e	dʒ	i	z	ʊ	v	o
z	ɛ	dʒ	e	z	ʊ	v	u
z	ɛ	dʒ	i	z	o	z	u
z	i	dʒ	e	z	u	z	o
z	I	dʒ	e	z	ʊ	z	o
z	i	dʒ	i	z	ʊ	z	u

Affricate/	Voiced	Unrounded	Front

Fricatives

Consonant 1	Vowel 1	Consonant 2	Vowel 2
z	I	dʒ	i
z	e	v	e
z	e	v	i
z	ɛ	v	e
z	ɛ	v	i
z	i	v	e
z	I	v	e
z	i	v	i
z	I	v	i
z	e	z	i
z	ɛ	z	e
z	ɛ	z	i
z	i	z	e
z	I	z	e
z	I	z	i

Affricate/ Fricatives	**Unvoiced**	**Unrounded**	**Front**	**Affricate/ Fricatives**	**Unvoiced**	**Rounded**	**Back**
Consonant 1	**Vowel 1**	**Consonant 2**	**Vowel 2**	**Consonant 1**	**Vowel 1**	**Consonant 2**	**Vowel 2**
f	e	f	i	f	o	f	u
f	ɛ	f	e	f	u	f	o
f	ɛ	f	i	f	ʊ	f	o
f	i	f	e	f	ʊ	f	u
f	I	f	e	f	o	s	o
f	I	f	i	f	o	s	u
f	e	s	e	f	u	s	o
f	e	s	i	f	u	s	u
f	ɛ	s	e	f	ʊ	s	o
f	ɛ	s	i	f	ʊ	s	u
f	i	s	e	f	o	tʃ	o
f	I	s	e	f	o	tʃ	u
f	i	s	i	f	u	tʃ	o
f	I	s	i	f	u	tʃ	u
f	e	tʃ	e	f	ʊ	tʃ	o
f	e	tʃ	i	f	ʊ	tʃ	u
f	ɛ	tʃ	e	s	o	f	o
f	ɛ	tʃ	i	s	o	f	u
f	i	tʃ	e	s	u	f	o
f	I	tʃ	e	s	u	f	u
Affricate/	**Unvoiced**	**Unrounded**	**Front**	**Affricate/**	**Unvoiced**	**Rounded**	**Back**

Fricatives

Consonant 1	Vowel 1	Consonant 2	Vowel 2
f	i	tʃ	i
f	I	tʃ	i
s	e	f	e
s	e	f	i
s	ɛ	f	e
s	ɛ	f	i
s	i	f	e
s	I	f	e
s	i	f	i
s	I	f	i
s	e	s	i
s	ɛ	s	e
s	ɛ	s	i
s	i	s	e
s	I	s	e
s	e	tʃ	e
s	e	tʃ	i
s	ɛ	tʃ	e
s	ɛ	tʃ	i
s	i	tʃ	e
s	I	tʃ	e
s	i	tʃ	i
s	I	tʃ	i
tʃ	e	f	e
tʃ	e	f	i
tʃ	ɛ	f	e
tʃ	ɛ	f	i
tʃ	i	f	e
tʃ	I	f	e
tʃ	i	f	i
tʃ	I	f	i
tʃ	e	s	e
tʃ	e	s	i
tʃ	ɛ	s	e
tʃ	ɛ	s	i
tʃ	i	s	e
tʃ	I	s	e
tʃ	i	s	i
tʃ	I	s	i
tʃ	e	tʃ	I

Fricatives

Consonant 1	Vowel 1	Consonant 2	Vowel 2
s	ʊ	f	o
s	ʊ	f	u
s	o	s	u
s	u	s	o
s	ʊ	s	o
s	ʊ	s	u
s	o	tʃ	o
s	o	tʃ	u
s	u	tʃ	o
s	u	tʃ	u
s	ʊ	tʃ	o
s	ʊ	tʃ	u
tʃ	o	f	o
tʃ	o	f	u
tʃ	u	f	o
tʃ	u	f	u
tʃ	ʊ	f	o
tʃ	ʊ	f	u
tʃ	o	s	o
tʃ	o	s	u
tʃ	u	s	o
tʃ	u	s	u
tʃ	ʊ	s	o
tʃ	ʊ	s	u
tʃ	o	tʃ	u
tʃ	u	tʃ	o
tʃ	ʊ	tʃ	o
tʃ	ʊ	tʃ	u

Affricate/ Unvoiced Unrounded Front

Fricatives

Consonant 1	Vowel 1	Consonant 2	Vowel 2
tʃ	ε	tʃ	e
tʃ	ε	tʃ	i
tʃ	i	tʃ	e
tʃ	I	tʃ	e
tʃ	I	tʃ	i