

**A Study on Sound Symbolism:
The Cognitive and Neural Mechanism Underlying
the Sound-meaning Correspondence in Language**

by

Sachi Itagaki

**Doshisha University
Graduate School of Life and Medical Sciences**

November 2019

Doshisha University
Graduate School of Life and Medical Sciences

Abstract

A Study on Sound Symbolism:
The Cognitive and Neural Mechanism Underlying
the Sound-meaning Correspondence in Language

by Sachi Itagaki

The correspondence between sound and meaning plays an important role in the language development. One idea that this relationship is not arbitrary is the sound symbolism that the sound itself has a specific impression. The purpose of this study is to elucidate the mechanism of sound symbolism from cognitive and neural perspective and whether sound symbolism affects the language development. In this study, a series of five experiments were conducted. 1) When word-like phonemes were used as sound stimuli, the sound symbolism related to visual size judgment was measured with functional magnetic resonance imaging (fMRI) scanning. The bilateral anterior cingulate cortex, left middle temporal gyrus, and right superior temporal gyrus were more activated under incongruent condition in sound symbolism. The results suggests that brain regions related to phoneme perception involved in sound symbolism. 2) Behavioral psychological experiment verified that phoneme perception affects sound symbolism. The results showed that the sound symbolism was difficult to observe if the sound stimulus was not clearly perceived as a single phoneme. This result suggested that not only acoustic features but also phoneme perception was important for sound symbolism. 3) The relationship between sound symbolism and phoneme perception was investigated with fMRI scanning. Data showed, that left postcentral gyrus was activated both in incongruent condition, and in phoneme perception. These results suggest that the motor cortex associated with tongue movement was activated even when sound stimuli were presented aurally, and that motor cortex involved in sound symbolism. The sound symbolism during the language development were measured using 4) infants' preferential looking behavior and 5) infants' preferential reaching behavior methods. The tone bursts, Japanese phonemes, and words were used as sound stimuli. The data showed that sound symbolism for tone bursts was acquired before the age of 9 months, and sound symbolism for Japanese phonemes was changed at approximately 12 months old, but sound symbolism in words was not exhibited at 15 months of age. These results suggested that the age to acquire the sound symbolism differs depending on the sound types. In all, the results suggest that motor cortex involved in phoneme articulation contributes to sound symbolism and that effect of sound symbolism and language develop through mutually interacting with each other. The findings in this study provide new insights into research aimed at elucidating the mechanisms of language development and language acquisition.

TABLE OF CONTENTS

LIST OF FIGURES.....vi

LIST OF TABLES.....vii

ACKNOWLEDGEMENTS.....viii

CHAPTER 1. Introduction

1.1 Sound symbolism.....1
1.2 Types of sound symbolism.....1
 1.2.1 Words and phonemes.....1
 1.2.2 Pitch of sound.....2
 1.2.3 Loudness of sound.....2
1.3 Mechanisms of sound symbolism.....2
 1.3.1 Acoustic character.....2
 1.3.2 Articulatory movement.....3
1.4 Motor Theory.....4
1.5 Language learning and sound symbolism.....5
 1.5.1 The role of sound symbolism in language.....5
 1.5.2 Language development.....5
1.6 Purpose.....6
1.7 Organization.....7

CHAPTER 2. Brain activity related to sound symbolism: Cross-modal effect of an aurally presented phoneme on judgment of size

2.1 Introduction.....8
2.2 Materials and Methods.....10
 2.2.1 Subjects.....10
 2.2.2 Experimental apparatus.....10
 2.2.3 Visual stimuli.....10
 2.2.4 Sound stimuli.....11
 2.2.5 fMRI parameters.....11
 2.2.6 Procedure.....12
 2.2.7 Analysis.....12
2.3 Results.....14

2.4 Discussion.....	16
2.5 Summary.....	18

CHAPTER 3. Sound symbolism on synthetic speech continuum in judgment of size of visual stimulus

3.1 Introduction.....	25
3.2 Materials and Methods.....	27
3.2.1 Subjects.....	27
3.2.2 Visual stimuli.....	27
3.2.3 Sound stimuli.....	27
3.2.4 Procedure.....	28
3.3 Results.....	29
3.3.1 Experiment 1: stimulus continuum /pi/-/ba/.....	29
3.3.2 Experiment 2: stimulus continuum /i/-/a/.....	30
3.4 Discussion.....	31
3.5 Summary.....	33

CHAPTER 4. Brain regions involved in sound symbolism and their relationship with phoneme perception

4.1 Introduction.....	37
4.2 Materials and Methods.....	40
4.2.1 Subjects.....	40
4.2.2 Experimental apparatus.....	40
4.2.3 Visual stimuli.....	40
4.2.3.1 Experiment 1.....	40
4.2.3.2 Experiment 2.....	41
4.2.4 Sound stimuli.....	41
4.2.5 fMRI parameters.....	42
4.2.6 Procedure.....	42
4.2.6.1 Experiment 1.....	42
4.2.6.2 Experiment 2.....	43
4.2.7 Analysis.....	43
4.2.8 Physiological noise removal.....	44
4.3 Results.....	46
4.4 Discussion.....	48
4.5 Summary.....	51

CHAPTER 5. Change of sound symbolism during the language development: Infants' preferential looking behavior

5.1 Introduction.....	58
5.2 Materials and Methods.....	61
5.2.1 Subjects.....	61
5.2.2 Experimental apparatus.....	61
5.2.3 Visual stimuli and sound stimuli.....	61
5.2.3.1 Experiment 1.....	61
5.2.3.2 Experiment 2.....	62
5.2.4 Procedure.....	63
5.2.5 Analysis.....	64
5.3 Results.....	65
5.4 Discussion.....	67
5.5 Summary.....	71

CHAPTER 6. Change of sound symbolism during the language development: Infants' preferential reaching behavior

6.1 Introduction.....	77
6.2 Materials and Methods.....	79
6.2.1 Subjects.....	79
6.2.2 Experimental apparatus.....	79
6.2.3 Object stimuli.....	79
6.2.4 Sound stimuli.....	80
6.2.5 Procedure.....	80
6.2.6 Analysis.....	80
6.3 Results.....	81
6.4 Discussion.....	82
6.5 Summary.....	83

CHAPTER 7. Conclusions

7.1 Summaries of Main Results.....	86
7.1.1 Brain activity related to sound symbolism: Cross-modal effect of an aurally presented phoneme on judgment of size (Chapter 2)	86
7.1.2 Sound symbolism on synthetic speech continuum in judgment of size of visual stimulus (Chapter 3)	88
7.1.3 Brain regions involved in sound symbolism and their relationship with phoneme	

perception (Chapter 4)	89
7.1.4 Change of sound symbolism during the language development: Infants' preferential looking behavior (Chapter 5)	90
7.1.5 Change of sound symbolism during the language development: Infants' preferential reaching behavior (Chapter 6)	91
7.2 Future works.....	92
7.2.1 The effect of sound symbolism on language learning.....	92
7.2.2 Changes in sound symbolism.....	93
7.3 Final remarks.....	93
REFERENCES	96
CURRICULUM VITAE	106

LIST OF FIGURES

Figure 2.1 Illustration of the experimental design and flow chart illustrating the experimental process.....	21
Figure 2.2 Average the RTs under each circle-size condition (Z-score).	22
Figure 2.3 Activation maps under the incongruent condition and under the $\pm 20\%$ incongruent condition.....	23
Figure 2.4 Contrast estimate for each target size in the right STG and the left MTG.....	24
Figure 3.1 Illustration of the experimental design.....	34
Figure 3.2 Average the RTs under each condition in both experiment 1 and 2 (Z-score)...	35
Figure 3.3 The differences in the Z scores for the RTs for all the sound stimuli.....	36
Figure 4.1 Illustration and flow chart of the experimental design in Experiment 1 and 2	53
Figure 4.2 Average the RTs for each condition (Z-score).....	54
Figure 4.3 Activation maps in the incongruent condition.....	55
Figure 4.4 Contrast estimate for each condition in each ROI related to the left/right precentral and postcentral gyri.....	56
Figure 4.5 Contrast estimate for each condition in the visual cortex during Experiment 1.....	57
Figure 5.1 Illustration of the experimental design in Experiment 1 and Experiment 2.....	72
Figure 5.2 The differences in looking time in each condition.....	73
Figure 5.3 The looking time in each condition in the shape condition.....	74
Figure 5.4 The looking time in each condition in the size condition.....	75
Figure 5.5 The differences in looking time in each condition.....	76
Figure 6.1 Example of two differently-shaped toys used as stimuli.....	85

LIST OF TABLES

Table 2.1 MRI activation clusters derived from comparison of the incongruent and congruent conditions.....	19
Table 2.2 MRI activation clusters derived from the comparison of the $\pm 20\%$ incongruent condition and the $\pm 20\%$ congruent condition.....	20
Table 4.1 MRI activation clusters related to motor area derived speech perception.....	52
Table 6.1 The first touch in infants' reaching behavior to objects in each subject.....	84

ACKNOWLEDGMENTS

I am most grateful to many people that have been supporting my study, and I wish thank all of them. Without their help, this dissertation would not have materialized.

First, I would like to express the deepest appreciation to Professor Shizuko Hiryu for providing me this invaluable opportunity to study as a doctoral student, for her constructive comments, and for encouraging me with her grateful support throughout years. I would also like to deeply thank Associate Professor Kohta I. Kobayashi for his insightful advices, warm encouragements, and many discussions that have enriched my research and my study life. I wish to thank Associate Professor Yoshitaka Matsuda for his grateful supports and shrewd advices to my research, and warm encouragements. I would like to express my gratitude to them for providing such a great research opportunity for my thesis work.

I would like to thank my collaborators Professor Yukuo Konishi, Professor Shoji Itakura, Associate Professor Masaharu Kato, and Dr. Toshihiko Aso for their great support, their advice, and constructive comments in carrying out the infant research. I wish to thank Mrs. Chikako Hashizume and Mrs. Mika Nishi and everybody at the Center for Baby Science for their work that recruiting infants and their warm encouragements. I am deeply grateful to the many students and more than 100 infants and their parents who participated in this study.

Many thanks to go to my collaborators on sound symbolism projects for adults; Mr. Shota Murai, Ms. Hikari Ohnishi, Mr. Hayao Kimura, and Mr. Makoto Matsumoto for insightful discussion and their patient effort and great support on projects. Many thanks to go to my collaborators on sound symbolism projects for infants; Ms. Momoko Hishitani, Ms. Haruka Suzuki, and Mr. Yuki Ito for valuable discussion and great support on infants' projects. I would like to express my sincere gratitude to Dr. Takafumi Furuyama, Dr. Hidetaka Yashiro, and Dr. Miwa Sumiya for their shrewd advices and warm encouragement, and Mr. Yuta Tamai, Ms. Momoka Nishimura and everybody at the laboratory for Neuroethology and Bioengineering Laboratory in Doshisha University for their support.

I would also like to thank Japan Society for the Promotion of Science (JSPS) for Grant-in-Aid for JSPS Fellows they provided, which made it possible for me to complete my thesis.

Finally, I would like to thank my parents: Hiroaki and Makie for their warm supports.

Sachi Itagaki

CHAPTER 1.

Introduction

1.1 Sound symbolism

The phenomenon of “sound symbolism” or “phonetic symbolism” in the association between meanings and sounds (i.e., phonemes) has been supported by several studies (e.g. Sapir, 1929; Köhler, 1947; Ramachandran and Hubbard, 2001). The idea underlying this phenomenon is that a sound itself conveys a certain impression, which then serves as the psychological basis for the word–meaning association.

1.2 Types of sound symbolism

1.2.1 Words and phonemes

Famous examples of sound symbolism are the “maluma-takete” and “bouba-kiki” effects (Köhler, 1947; Ramachandran and Hubbard, 2001). These studies have shown that the pseudo words “maluma” and “bouba” are associated with a rounded shape, while the words “takete” and “kiki” are associated with a spiky shape. In addition to sound symbolism in pseudo words, sound symbolism is also observed in phonemes or ideophones. In one study, it was observed that the phoneme /a/ is “larger-sounding”, while the phoneme /i/ is “small-sounding” (Ohtake and Haryu, 2013). Knoeferle et al. (2017) reported that voiceless stop consonants are considered spiky and smaller, while nasal consonants are considered rounded and larger. Meanwhile, Lockwood et al. (2016) used unknown natural language ideophones to investigate whether naïve subjects could identify the correct translation of the word. According to the results, subjects were able to identify the correct translation of the unknown ideophone. These studies revealed that sound symbolism could accommodate sound-to-meaning mapping of words and phonemes.

1.2.2 Pitch of sound

Sound symbolism has also been observed for pitch (Bernstein and Edelstein, 1971; Chiou and Rich, 2012). In one study, the interaction between pitch and brightness was measured (Marks, 1987). The subjects' responses to dark and dim visual stimuli were faster and more accurate when the accompanying sound was low rather than high in pitch, while their responses to light and bright visual stimuli were faster when the accompanying sound was high rather than low in pitch. However, responses to low (or high) tones were faster when the visual stimulus was dim/dark (or bright/light). This correspondence between pitch and brightness was observed even in chimpanzees (*Pan troglodytes*; Ludwig et al., 2011). Other studies have suggested that high pitch sounds are associated with the impression of smallness, while low pitch sounds are associated with the impression of largeness (Gallace and Spence, 2006; Walker and Smith, 1985; Parise and Spence, 2008). Knoeferle et al. (2017) stated that when the higher first formant (F1) is paired with longer vowel duration, it was associated with the impression of a larger size. Surprisingly, the lower second formant (F2) was also associated with the impression of a larger size. In other words, the impression of largeness or smallness may be elicited to a vowel's F2 frequencies. These studies confirmed that pitch can influence the impression of size.

1.2.3 Loudness of sound

In addition, sound symbolism for loudness has also been investigated (Marks, 1974, see Spence, 2011, for a review). One study demonstrated that both adults and children (5-year-olds) reliably matched brightness to loudness crossmodally (Bond and Stevens, 1969). In other words, both groups of participants (adults and children) paired a light gray patch of color with a loud sound and a dark gray patch of color with a quiet sound. These results suggest that even with loudness, some impression is made.

1.3 Mechanisms of sound symbolism

1.3.1 Acoustic character

The correspondence between pitch and size may reflect the characteristics of

acoustic resonance. In terms of the size of a human body, a larger body resonates at a lower frequency than a smaller body. Once again, it could be argued that such correspondence might depend on the physical acoustic properties. For instance, harder objects tend to resonate at higher frequencies and break into sharper pieces than softer objects (e.g. Freed, 1990; Walker et al., 2010). This probably explains the cross-modal correspondence between pitch and object shape (Walker et al., 2010). In addition, large objects that normally resonate at low frequencies are generally considered heavy. Therefore, it is unlikely to fly or be found at a higher position; this may explain the correspondence between pitch and altitude. These ideas perhaps explain the cross-modal correspondence of shape and size based on the formant frequency for sound symbolism in phonemes.

1.3.2 Articulatory movement

Some studies show that articulatory movement also contributes to sound symbolism (Newman, 1933; Sapir, 1929; Shinohara and Kawahara, 2010). For example, vowels such as /a/, /u/, and /o/ convey the impression of a large-sized object, whereas vowels such as /i/ or /e/ evoke the impression of a small-sized object. When we pronounce these vowels, the position of the tongue and the oral cavity differ. These differences may contribute to a vowel-size correspondence. On the other hand, when we pronounce the Japanese phoneme /a/, the position of the tongue is toward the back of the mouth and the oral cavity is larger. On the other hand, when we pronounce the Japanese /i/, the position of the tongue is toward the front of the mouth, hits the upper jaw, and the oral cavity is smaller. One important theory to consider is the motor theory, which suggests that speech pronunciation is related to speech recognition (Lieberman and Whalen, 2000). One study observed motor-evoked potentials related to anterior tongue muscles while participants were listening to words (Fadiga et al., 2002). In order to clarify whether articulatory movement contributes to sound symbolism, the relationship of the neural basis between articulatory movement and sound symbolism should be examined in addition to the aforementioned behavioral experiments. For example, in designing an experiment that directly examines the relationship between articulatory movement and sound symbolism,

a behavioral experiment may ask the subject to verbally respond at the time sound symbolism is observed. However, this experiment is not enough to clarify whether unconscious movement, which could have occurred while listening to the voice, contributes to sound symbolism. Therefore, the relationship between motor areas and sound symbolism should be verified through experiments that measure brain activity.

1.4 Motor Theory

Humans are not only speakers but also listeners. Based on this fact, motor theory suggests that the process of speech production plays an important role in speech perception. Among the basic claims of motor theory that have been modified in response to various arguments are the following (Liberman and Mattingly, 1985):

- Speech perception is based on estimating the articulatory gesture of the speaker who generated it from the acoustic signal.
- Estimation of articulatory gestures is based on “analysis by synthesis” using an internal model of articulatory organs.

The McGurk effect is an experiment that examines the contribution of a speaker’s articulatory gestures to speech perception (McGurk and MacDonald, 1976). When the sound /ba/ and the video of a person saying /ga/ are synchronized, the sound is perceived as /da/ or /ga/. The sound of /ba/ is not articulated unless the mouth closes, but the mouth does not close in the /ga/ video. Therefore, it is not perceived as /ba/ but is perceived as another sound, /da/. In this study, it was suggested that speech perception was influenced by visual information and was estimated based on the articulatory gestures of the speaker. Motor theory has also been investigated for its neural basis, and it has become clear that multiple parallel circuits are involved in speech processing (Scott and Johnsrude, 2003; Hickok and Poeppel, 2000). One auditory circuit takes a ventral route that travels forward from the primary auditory cortex and passes through the anterior part of the superior temporal sulcus to the ventral region of the frontal lobe. The other auditory circuit takes a dorsal pathway that travels backward from the primary auditory cortex and passes through the posterior part of the superior temporal sulcus and the parietal association area

to each region of the frontal lobe such as the premotor cortex. In order to investigate the path of speech perception that contributes to sound symbolism, the brain activity that contributes to sound symbolism should be measured.

1.5 Language learning and sound symbolism

1.5.1 The role of sound symbolism in language

Language is what sets mankind apart. In other words, humans communicate using language, making our communicative behavior different from that of other animals. Animals also communicate by voice, such as calls, but humans have far more means to use voice-based communication. Language has several constituents such as words and grammar. The use of words, i.e., the correspondence between an object and the word that describes that object, has been considered arbitrary in natural languages (de Saussure, 1983). For example, “dog” refers to a carnivorous quadruped mammal that is raised as a pet, to assist police, and for other purpose. If the correspondence between words and objects is arbitrary, the name “cat” could be used instead of “dog”. Since word-to-meaning mapping has this quality of arbitrariness, it is considered that humans are able to communicate infinitely, unlike animals. However, language learning would be more difficult if word-to-meaning mapping was completely arbitrary. Practically, there may be rules, such as sound symbolism. Conventions such as sound symbolism could contribute to word-to-meaning mapping. In addition, thanks to these rules, we may be able to perform semantic processing with limited hesitation when encountering an unknown word (Lockwood et al., 2016).

1.5.2 Language development

There are several phases in language development. First, infants learn to produce non-speech sounds. Next, the infant begins to produce vowel-like sounds and canonical babbling. Then, language-specific speech production begin at approximately 10 months of age, with the first word often produced at approximately 12 months of age (Kuhl, 2004). As the number of words infants can produce increases, they will be able to speak sentences with two or three words. Vocalization development in infants begins with a

language-universal pattern of production that eventually becomes language-specific (Kent, 1992; Oller and Lynch, 1992; Stoel-Gammon, 1992). Two elements of anatomical change and language learning are important in the early phase of infant utterance. The vocal tract of young infants is significantly different from that of adults and more closely resembles that of non-human primates than that of adult humans (Bosma, 1975; Kent, 1981). An infant's vocal tract is not only much smaller than that of an adult's, but also has a broader oral cavity, a tongue mass that is proportionally larger and more anterior, and a more gradually inclined oropharyngeal tract (see Kent, 1992, for a review). Infants later develop a vocal tract similar to that of adults. This development increases the range of formants that can be produced. In one study, it was observed that changes in formant production occur during infancy (Kuhl and Meltzoff, 1996). In 12-week-old infants, the distribution of vowels' formants was compact and overlapped. In 20-week-old infants, the distribution of vowels' formants had increased and the overlap between vowels was much reduced. There are several phases in speech perception (Kuhl, 2004). Shortly after birth, infants can discriminate between phonetic contrasts of all languages; language-specific perception for vowels occurs at approximately 6 months of age. At approximately 12 months of age, infants can recognize language-specific sound combinations and demonstrate an increased perception of native-language consonants. Sound symbolism may interact with the recognition of sound combinations and the production of formants.

1.6 Purpose

Although these sound symbolism studies suggested that both phoneme perception and simple acoustic characteristics are essential (Newman, 1933; Gallace and Spence, 2006), the relative contribution of each factor and the interactions between them have not been investigated systematically. I think it is necessary to investigate the effects of these factors on sound symbolism, respectively. The purpose of this study is to elucidate the mechanism of sound symbolism from the cognitive and neural perspectives and whether sound symbolism affects language development. Therefore, I first constructed an experimental paradigm that can measure sound symbols in an unusual environment such as functional magnetic resonance imaging (fMRI), and measured brain activity that

contributes to sound symbolism. Next, in order to investigate the relationship between phoneme perception and sound symbolism, which was suggested to be an essential factor, I used behavioral, psychological experiments to measure the effect of phoneme perception difficulty on sound symbolism. Furthermore, it was investigated whether brain regions related to phoneme perception contribute to sound symbolism. Like previous studies, sound symbols can contribute to language acquisition (Lockwood et al., 2016; Imai et al., 2015). Thus, in this study, the sound symbolism during the language development was also examined by using the preferential looking behavior and preferential reaching behavior methods of infants.

1.7 Organization

This dissertation is organized as follows. Chapter 2 contains brain activity related sound symbolism to word-like phonemes. Chapter 3 describes the relationship between phoneme perception and sound symbolism during behavioral psychological experiment. Chapter 4 presents the neural basis between phoneme perception and sound symbolism. In Chapter 5, sound symbolism for infants was examined using infants' preferential looking behavior method. Chapter 6 presents sound symbolism for infants was measured with infants' preferential reaching behavior method. Finally, conclusion and possible directions for future research are described in Chapter 7.

CHAPTER 2.

Brain activity related to sound symbolism: Cross-modal effect of an aurally presented phoneme on judgment of size

2.1 Introduction

It has been assumed that in natural language, there is no particular regularity in the correspondence between an object and the acoustic properties of the word that describes that object (de Saussure, 1929). However, a phenomenon called “sound symbolism” or “phonetic symbolism” in the association between meanings and sounds (i.e., phonemes) has been confirmed by several studies (Köhler, 1947; Auracher et al., 2010; Shinohara et al., 2016; Ludwig et al., 2011; Hirata et al., 2011). The idea underlying this phenomenon is that a sound itself makes a certain impression, which then serves as the psychological basis for the word–meaning association.

The most famous example of this phenomenon is the so-called “bouba-kiki” effect (Ramachandran and Hubbard, 2001), which involves asking a participant to name spiky and round shapes using only the sounds “bouba” and “kiki.” According to the results, most people associate “bouba” with “round” and “kiki” with “spiky.” Moreover, this effect has been observed regardless of the subject’s age or native language. Most sound symbolism studies have used a relatively naturalistic approach (Parise and Pavani, 2011; Parise and Spence, 2012; Knoeferle et al., 2017; Newman, 1933). In one study, in an attempt to illustrate the effect of each phoneme, subjects were asked about their impressions of an artificial word (Knoeferle et al., 2017). These studies successfully detected a relationship between the acoustic features of a stimulus word and the impression left by or the response to these features. However, the neural basis of sound symbolism is less well understood.

Kovic and colleagues analyzed the neural processes underlying novel word–visual object matching using electroencephalography (EEG) (Kovic et al., 2010). Their results

showed that sound–symbol matching evoked an early negative EEG response, indicating that such matching involved early sensory processes. Recent EEG research with an 11-month-old infant showed that large-scale synchronization within the left hemisphere was sensitive to the sound–symbol correspondence (Asano et al., 2015). These results suggest that the preverbal infant is capable of mapping an auditory stimulus onto a visual experience by recruiting a multimodal perceptual processing system. One early neuroimaging study of sound symbolism (Japanese onomatopoeia) was conducted by Osaka and colleagues, who used fMRI to investigate how onomatopoeic words modified the activity of different cerebral regions (Osaka et al., 2003; Osaka et al., 2004). Recently, Kanero et al. studied the neural basis of sound symbolism with fMRI using the written text of Japanese mimetic words and moving images (Kanero et al., 2014). Their results showed that the right temporal region was involved in processing Japanese mimetic words. Here, I used the sounds of phonemes as auditory stimuli, and examined how different aspects of sound symbolism (phoneme vs mimetic word) and different modalities of stimulus presentation (visual vs auditory) affect the activities of the right temporal and other brain regions.

In this study, I examined the effect of a phoneme (the basic unit of sound within a word) on the judgment of the size of a visual stimulus and the related brain activity. Subjects were required to perform two types of task, a comparison task and a control task (Fig. 2.1). In the comparison task, they compared the visual sizes of standard and target stimuli. The target stimulus was either smaller or larger than the standard by $\pm 5\%$, $\pm 10\%$, or $\pm 20\%$ of its diameter, and was displayed to the subjects while they were listening to the syllable “bobo” or “pipi”. Previous research into sound symbolism predicted a tendency to pair the former sound with a larger object and the latter with a smaller object (Newman, 1933). The control task was designed to prevent subjects ignoring the sound stimuli.

2.2 Materials and Methods

2.2.1 Subjects

Fourteen subjects (four females and ten males; aged 21–26 years) participated in the fMRI experiment after they had provided their written informed consent. All the subjects were right-handed native Japanese speakers. None of the participants had any knowledge of sound symbolism or the experiment. The data from three participants were excluded because of artifact or inadequate task performance (e.g., head movement > 3 mm). The experimental protocol was approved by the Research Ethics Committee on human subjects of Doshisha University, and the study carried out in accordance with the guidelines of the committee.

2.2.2 Experimental apparatus

Each subject was positioned supine in an MRI scanner. The sound stimuli were presented through MRI-compatible headphones (Kiyohara Optics Inc., Tokyo, Japan), and the visual stimuli were presented with a projector and mirror system. The subjects viewed the visual stimuli projected onto a mirror placed 18 cm in front of their eyes, and they pushed buttons (Current Designs, Inc., PA, USA) with their left hand to respond. Experiment control software (Presentation®; Neurobehavioral Systems, Inc., Albany, CA, USA) was used to synchronize the experimental procedures with the fMRI scans.

2.2.3 Visual stimuli

In this experiment, I examined the effect of sound symbolism when judging the size of a visual stimulus. The visual stimulus, a gray circle that looked like a doughnut, was presented on a frosted screen at the end of the scanner bore, and could be seen by the subject via a mirror mounted on the head coil. The standard stimulus had an outer circle of 300 pixels and an inner circle of 280 pixels. The target stimulus was either smaller or larger than the standard stimulus by $\pm 5\%$, $\pm 10\%$, or $\pm 20\%$ of its diameter. In total, seven sizes (one for the standard and six for the targets) were used. Each stimulus was presented twice for 200 ms, with an interstimulus interval (ISI) of 120 ms (Fig. 2.1A). A red cross (34 pixels) was always presented as the fixation point at the center of the screen.

2.2.4 Sound stimuli

The sound stimuli were “bobo” and “pipi”. A publicly available sound dataset (FW03; NTT Communication Science Laboratories, Kanagawa, Japan) was used to create the sounds. All sounds were recorded at a sampling frequency of 48 kHz and a quantization of 16 bits. The single-syllable utterances /bo/ and /pi/ were spoken by a male, and these were duplicated to produce the sound stimuli “bobo” and “pipi”, respectively. These sound stimuli have no accent, therefore they were not similar to any word. The duration of the sound was 520 ms, and the stimulus amplitude was 64 dB sound pressure level (SPL). According to previous research, the sound /pi/ was louder than /bo/ by about 1.4 dB in subjective loudness (Nagatani et al., 2008). The sound stimulus was synchronized with the visual stimulus. In an additional experiment, the participants were asked to evaluate the size of each sound stimulus by selecting labeled pictures of five Russian nested dolls of different sizes (doll 5 was the largest) (Fort et al., 2015; Knoeferle et al., 2017). The mean scores for “bobo” and “pipi” were 4.0 and 2.3, respectively, and no participant ever scored “bobo” smaller than “pipi”. Therefore, I confirmed that “bobo” created a larger impression than “pipi”.

2.2.5 fMRI parameters

Functional images of brain activity were acquired with a 1.5-T MRI system (Echelon Vega, Hitachi Medical Corporation, Tokyo, Japan) as T2*-weighted images using a gradient echo–echo planar imaging (GE–EPI) sequence with a resolution of $3 \times 3 \times 5$ mm voxels (30 axial slices; field of view [FOV]: 192 mm; matrix: 64×64 ; repetition time [TR]: 3000 ms; echo time [TE]: 50 ms; flip angle [FA]: 90°). The first five scans were discarded to avoid magnetic saturation effects. A structural T1 image was acquired at a resolution of $1 \times 1 \times 1$ mm using a three-dimensional gradient echo inversion recovery (3D-GEIR) sequence (192 slices, 1 mm thick; sagittal; FOV: 256 mm; matrix: 256×256 ; TR: 9.7 ms; TE: 4 ms; T1: 1045 ms; FA: 8°).

2.2.6 Procedure

The subjects were asked to judge the difference in the sizes of the standard and target stimuli (Gallace and Spence, 2006). Each trial began with a 3,000 ms rest period. The standard stimulus was then presented for 520 ms, followed by a 300 ms ISI. The target stimulus was then presented for 520 ms, followed by a response period of 3,000 ms (Fig. 2.1A). The screen was black during the rest and response periods. After the subject had responded to the task by pressing a button, the next trial began automatically. When the sound (sound 2) presented with the target stimulus was identical to the sound (sound 1) presented with the standard stimulus, the subject had to press the middle button, regardless of the visual stimulus (control task). When the sound stimulus presented with the target stimulus differed from that presented with the standard stimulus, the subject had to respond according to whether the target circle was smaller or larger than the standard (Fig. 2.1B). The subjects used their index and ring fingers to press the left and right buttons, respectively (comparison task). The button assignment (i.e., which button represented which answer) was changed between subjects. There were six combinations of visual stimuli (one standard \times six targets) and four combinations of sound stimuli (two sound 1 \times two sound 2), yielding a total of 24 stimulus combinations. The entire stimulus set was randomized to create one block (24 trials), and each session consisted of three blocks. In the behavioral experiment, each subject completed one session (72 trials).

2.2.7 Analysis

The functional imaging data were processed with the SPM12 software (Wellcome Department of Cognitive Neurology, London, UK). For realignment, the image was spatially normalized to an EPI template in Montreal Neurological Institute (MNI) space, and smoothed with an 8 mm full width at half maximum Gaussian kernel. To examine the brain activation associated with phonetic symbolism, I created different types of contrast images (congruent minus incongruent for all target sizes combined, incongruent minus congruent for all target sizes combined, congruent minus incongruent for $\pm 20\%$ target sizes, and incongruent minus congruent for $\pm 20\%$ target sizes). These effects were modeled as stick functions convoluted with the canonical hemodynamic response

function. The movement parameters of the realignment corrections were included in the model as covariates of no interest. Activated areas were depicted on a standard human brain template from MRIcro software (www.mccauslandcenter.sc.edu/crnl/micro).

To further clarify the effect size, I used a ROI analysis of the contrast estimates in the right STG and left MTG. Each ROI was constructed with leave-one-subject-out cross-validation to avoid any bias arising from nonindependence (Esterman et al., 2010). Briefly, the data for each subject were iteratively left out of the group. The resulting group ($N - 1$ subjects) analyses returned peak coordinates of the ROIs (spheres with a 4 mm radius) for the subject who was left out, and the average responses in the ROIs in the right STG and left MTG were estimated.

2.3 Results

I analyzed the reaction times (RTs) for the compared tasks. Several studies have shown that the phonemes /b/, /d/, /g/, and /o/ elicit “larger” reactions, whereas /p/, /t/, /k/, and /i/ elicit “smaller” reactions (Newman, 1933; Sapir, 1929). Therefore, I assumed that “bobo” and “pipi” would induce “larger” and “smaller” impressions, respectively. I defined the congruent and incongruent conditions as follows and analyzed the data accordingly. Under the congruent condition, the target visual stimulus was consistent with the reaction to the sound (i.e., the larger target was presented with “bobo” or the smaller with “pipi”). Under the incongruent condition, the target visual stimulus was inconsistent with the reaction to the sound (i.e., the larger target was presented with “pipi” or the smaller one with “bobo”).

Eleven subjects performed well on the comparison task. The mean correct response rate was 94.1%. The differences in the RTs under the congruent and incongruent conditions are shown in Fig. 2.2. The mean RT decreased as the size difference between the targets increased from $\pm 5\%$ to $\pm 20\%$. The mean RT under the incongruent condition was longer than that under the congruent condition for all target sizes. However, this difference was statistically significant only when the target was $\pm 20\%$ of the standard ($t = -5.93, p < 0.001$, t test with Bonferroni correction).

In this study, I investigated the brain region associated with sound symbolism by contrasting the incongruent condition with the congruent. The bilateral anterior cingulate cortex (ACC) was activated under the incongruent minus congruent condition for all target sizes (Fig. 2.3A, Table 2.1)

The activation of the right lingual gyrus, left middle temporal gyrus (MTG), and right superior temporal gyrus (STG) was observed under the $\pm 20\%$ incongruent minus $\pm 20\%$ congruent condition (Fig. 2.3B, Table 2.2). The contrast (incongruent – congruent) for each target size condition in each of these brain regions is presented in Fig. 2.4. There was no significant activation under the $\pm 20\%$ congruent minus $\pm 20\%$ incongruent condition.

The contrast estimates (incongruent – congruent) for the right STG and left MTG under each condition were evaluated with a region-of-interest (ROI) analysis with leave-

one-subject-out cross-validation, and are shown in Fig. 2.4, in which a positive value indicates stronger activation during the incongruent condition than during the congruent condition. The difference at $\pm 20\%$ was more prominent than that at $\pm 5\%$ or $\pm 10\%$ in each region.

2.4 Discussion

The RT under the congruent condition was shorter than that under the incongruent condition, suggesting that the effect of sound symbolism was observed under my experimental setting (the phoneme as a sound stimulus). However, the efficacy of sound symbolism differed with the target size, and only the $\pm 20\%$ condition yielded a significant difference between the congruent and incongruent conditions. Because the RTs were more strongly affected by the target size than by the congruent–incongruent difference (Fig. 2.2), it is not surprising that the effect of sound symbolism was minimal under some target conditions.

The bilateral ACCs were more strongly activated under the incongruent condition than under the congruent condition. The activation of the dorsal ACC has been associated with cognitively demanding tasks, typically involving response conflict, such as the Stroop task (Pardo et al., 1990; Roelofs et al., 2006). In this study, I observed a more pronounced BOLD response in the dorsal ACC, suggesting that the task performed under the incongruent condition was more cognitively demanding for the subjects. Together with the longer RTs, the activation of the dorsal region of the ACC may reflect the Stroop-like interference between the phoneme and the visual magnitude of the stimulus.

The left MTG was more strongly activated under the $\pm 20\%$ incongruent condition than under the $\pm 20\%$ congruent condition. The results of the ROI analysis of the peak region are plotted in Fig. 2.4. The left MTG has been identified in previous studies as a brain region related to semantic association (see Price, 2012, for a review). This region is also very similar to the area that was activated under the incongruent condition in a priming study using EEG (Schneider et al., 2008), in which congruence was based on the relationship between a picture and an environmental sound (i.e., an animal and its vocalization). That study demonstrated that the left MTG is involved in cross-modal semantic-matching processes. The result of this study implies that sound symbolically matching between the target size and a phoneme is also processed in the MTG.

The right STG was more strongly activated under the $\pm 20\%$ incongruent condition than under the $\pm 20\%$ congruent condition. The ROI analysis of the peak region (Table 2.2) revealed that the activation of the STG (incongruent minus congruent) was more

prominent as the size difference between the targets increased from $\pm 5\%$ to $\pm 20\%$ (Fig. 2.4), indicating that the efficacy of the sound symbolism correlated with the activation of the region. Many previous studies have identified the STG as a primary region for speech perception (see Price, 2012, for a review). Interestingly, activities in the right STG have been associated with the incongruence between emotional prosodic cues and other information (speech content (Wittfoth et al., 2010) and facial expression (Watson et al., 2013)). The result of this study suggests that the right STG is part of a brain network involved in processing conflict in phonemic sound symbolism in addition to emotional prosodic information.

Kanero and colleagues (2014) reported that the right posterior STG was more strongly activated when subjects were evaluating matched pairs of mimetic words and moving images than when they were evaluating mismatched pairs. However, my analysis detected no significant activation in this region. Although the functional localization of the right STG is still contentious, several researchers have reported that the anterior region is strongly associated with phoneme perception (Obleser et al., 2006; Correia et al., 2015). The stimulus difference (mimetic word in Kanero's study or phoneme in my study) could contribute to the difference in the activated area. In addition to analyzing the stimulus difference, they asked their subjects to actively report their impression of mimetic words, whereas subjects in this study were engaged in a visual discrimination task, and were not asked to report their impression of the phoneme. The observed differences in activation area could be attributable to these task differences. Further research is required to investigate the sound symbolic effect on both phonemes and mimetic words within the same experimental paradigm to detect the functional segregation, if any, of the right STG into the different aspects of sound symbolism. There are several limitations of this study. The number of subjects is relatively small, and only two types of speech sounds were tested; therefore, the data of this study should be treated as a pilot study to draw general conclusion on the neural processing of sound symbolism. I still believe that the data of this study indicates the involvement of bilateral temporal regions (i.e., the right STG and left MTG) in the sound-meaning association.

2.5 Summary

Sound symbolism is the theory that a sound conveys a certain impression (e.g., the phoneme /p/ is associated with an impression of smallness) and could be the psychological basis of the word–meaning association. In this study, I investigated the neural basis of sound symbolism. Subjects were required to compare the visual sizes of standard and target stimuli while listening to syllables that were either large-sounding or small-sounding. Stimulus–response congruence was defined as the agreement between the target size and the syllable’s impression. Behavioral data showed that the subjects exhibited a longer reaction time in the incongruent condition than in the congruent condition, indicating that they tended to associate object size with certain syllables. I used fMRI to evaluate cerebral activity during the task, and found that both semantic- and phonetic-process-related areas of the brain (left middle temporal gyrus and right superior temporal gyrus, respectively) were activated in the incongruent condition. Furthermore, the bilateral ACCs were more strongly activated in the incongruent condition than in the congruent condition. The activation of the dorsal ACC has been associated with cognitively demanding tasks, typically involving response conflict, such as the Stroop task. These results suggest that these regions were associated with the incongruence of sound symbolism.

Brain region	MNI coordinates			Z-score	Cluster size (voxels)
	x	y	z		
Right anterior cingulate cortex	10	42	2	4.91	569
Left anterior cingulate cortex	0	32	8	4.78	

Table 2.1 MRI activation clusters derived from comparison of the incongruent and congruent conditions.

Note: Coordinates are in MNI (Montreal Neurological Institute) space. Included are clusters that survived the threshold at a voxel level of $p < 0.001$ with a cluster-level family-wise error (FWE) correction of $p < 0.05$.

Brain region	MNI coordinates			Z-score	Cluster size (voxels)
	x	y	z		
Right lingual gyrus	32	-60	-2	4.35	774
Right inferior occipital gyrus	40	-60	-2	4.14	
Right lingual gyrus	18	-40	-2	3.95	
Right inferior occipital gyrus	36	-62	0	3.89	
Right calcarine cortex	32	-64	6	3.88	
Left middle temporal gyrus	-48	-44	0	4.22	1040
Left middle occipital gyrus	-30	-76	8	4.06	
Left middle temporal gyrus	-48	-50	-2	4.05	
Left middle temporal gyrus	-44	-50	0	4.01	
Left lingual gyrus	-26	-54	-4	3.76	
Right frontal operculum	50	16	-6	4.00	551
Right temporal pole	44	10	-14	3.73	
Right posterior orbital gyrus	28	26	-10	3.73	
Right planum polare	44	-6	-10	3.71	
Right superior temporal gyrus	48	-10	-12	3.57	

Table 2.2 MRI activation clusters derived from the comparison of the $\pm 20\%$ incongruent condition and the $\pm 20\%$ congruent condition.

Note: Coordinates are in MNI space. Included are clusters that survived the threshold at a voxel level of $p < 0.005$ with a cluster-level FWE correction of $p < 0.05$.

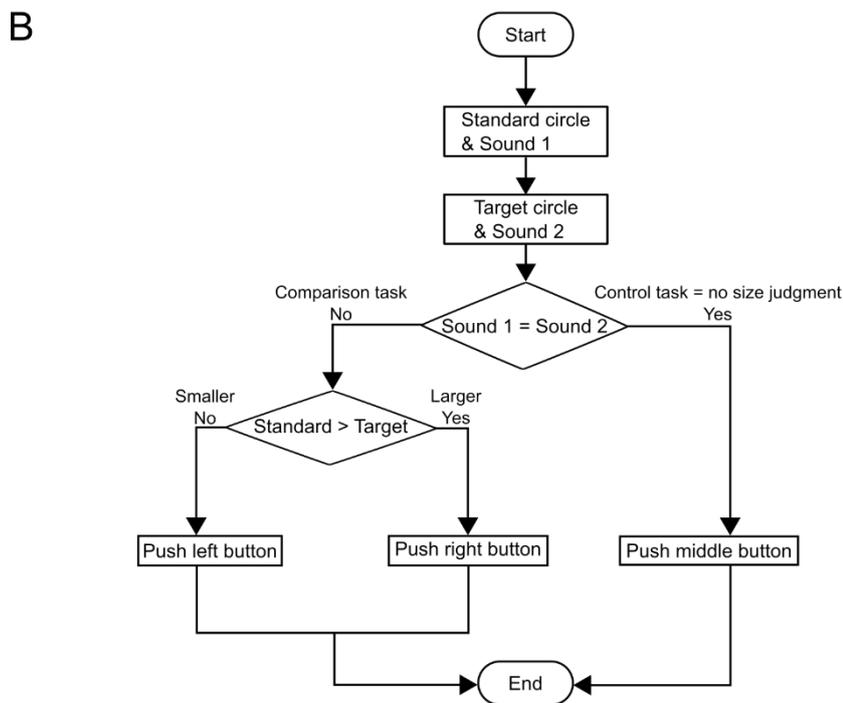
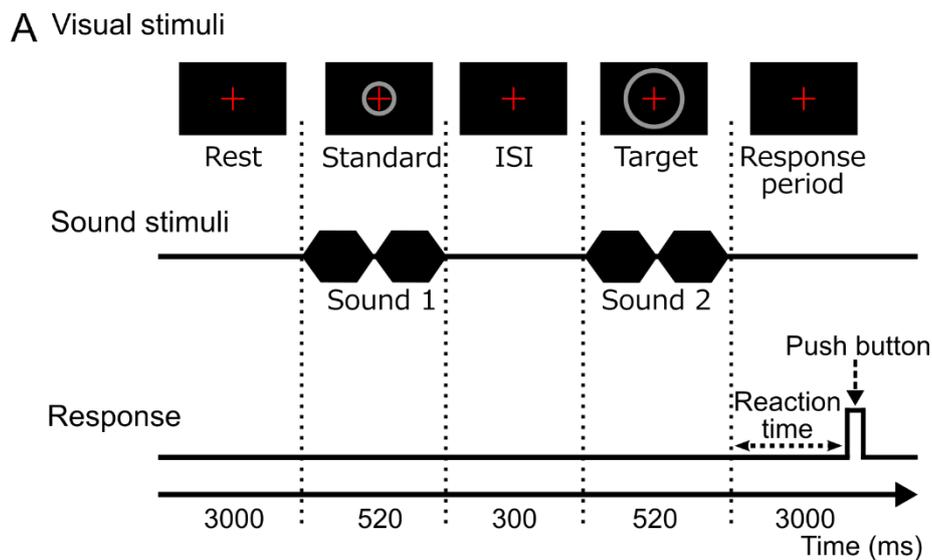


Figure 2.1 Illustration of the experimental design and flow chart illustrating the experimental process.

(A) Illustration of the experimental design. Each trial started with a rest period (3,000 ms). The standard stimulus (520 ms) was then presented, followed by the target stimulus (520 ms) after a 300 ms interstimulus interval (ISI). Subjects were instructed to respond to the task during the response period (3,000 ms) via a button device, using the fingers of their left hand. A red cross was always presented as the fixation point.

(B) Flow chart illustrating the experimental process. Subjects were instructed to push the middle button to respond to the control task and to push the left or right button to respond to the comparison task. Button assignment (which button represented a specific answer) in the comparison task was changed between subjects.

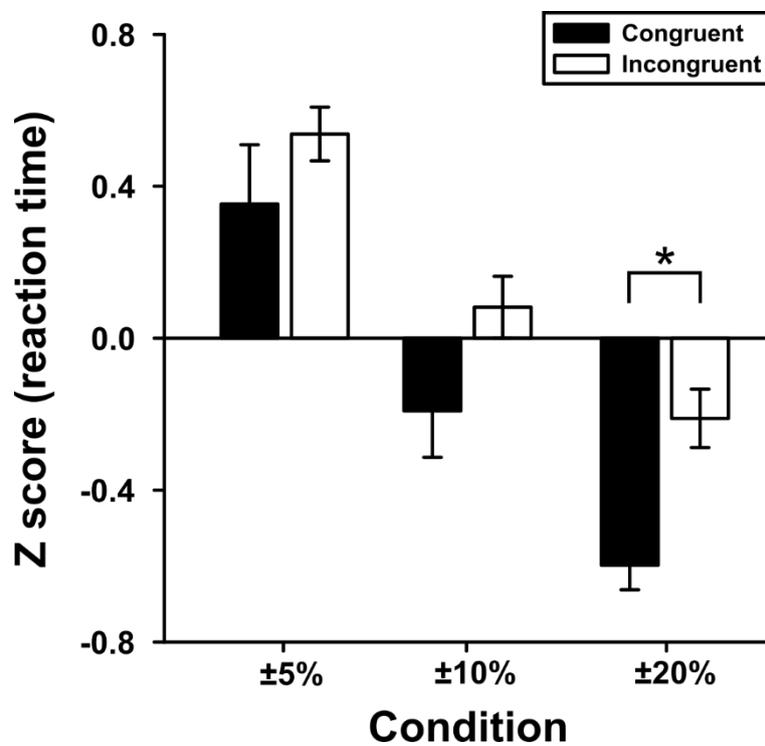


Figure 2.2 Average the RTs under each circle-size condition (Z-score). The vertical axis represents the Z-score and the horizontal axis represents the target size. Error bars represent the standard errors of the means. The Z-score was greater under the incongruent condition than under the congruent condition for all target-size conditions; the difference under the $\pm 20\%$ target-size condition was significant ($t = -5.93, p < 0.001$).

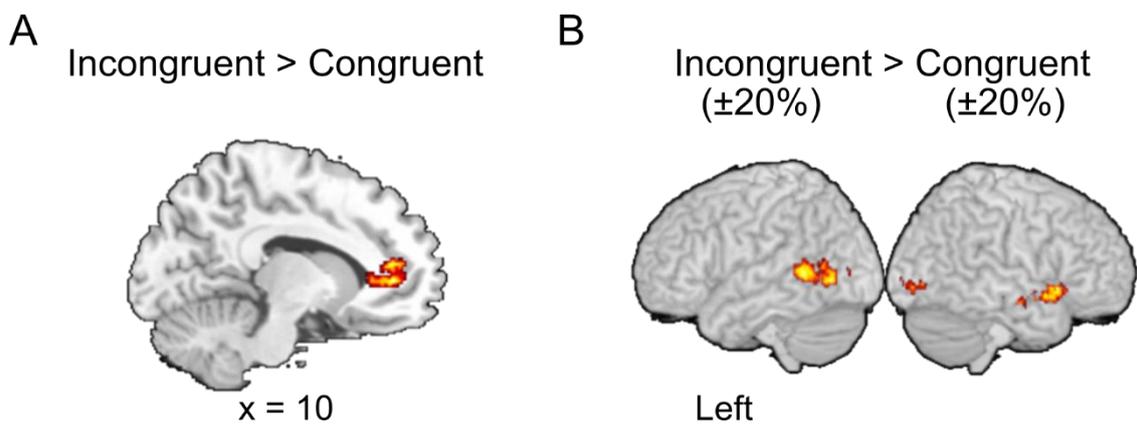
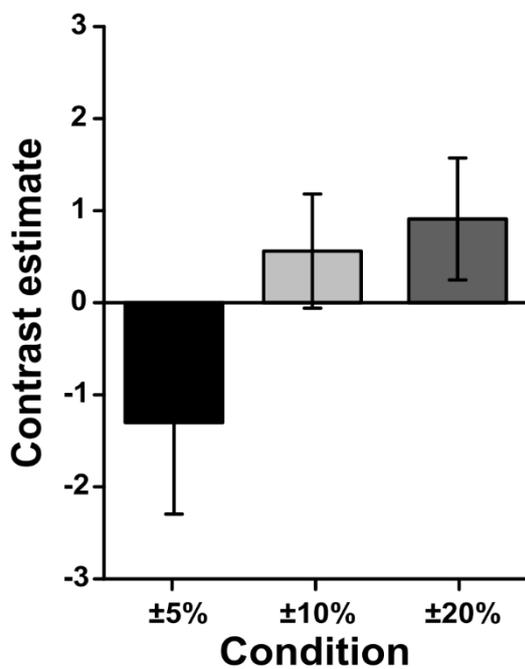


Figure 2.3 Activation maps under the incongruent condition and under the $\pm 20\%$ incongruent condition.

(A) Activation maps under the incongruent condition. The brain region showing greater activity under the incongruent condition than under the congruent condition is the ACC [$p < 0.001$ with a cluster-level family-wise error (FWE) correction of $p < 0.05$].

(B) Activation maps under the $\pm 20\%$ incongruent condition. Shown are the brain regions with greater activation during the $\pm 20\%$ incongruent condition than during the $\pm 20\%$ congruent condition ($p < 0.005$ with a cluster-level FWE correction of $p < 0.05$). Significant regions were superimposed on a standard brain template from MRICro software.

Right superior temporal gyrus



Left middle temporal gyrus

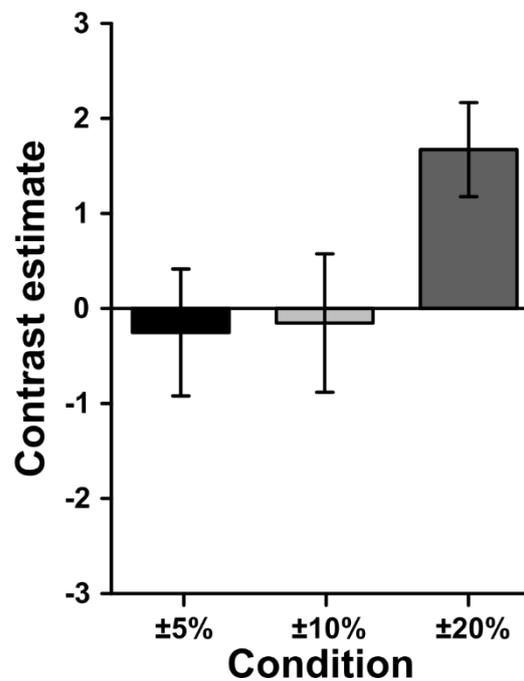


Figure 2.4 Contrast estimate for each target size in the right STG and the left MTG. The contrast estimate for each target size, calculated by subtracting the congruent from the incongruent condition, in the right superior temporal gyrus and left middle temporal gyrus. The ROIs were spheres with a radius of 4 mm. Error bars represent the standard errors of the means. The contrast estimate was greater for a difference of $\pm 20\%$ than for a difference of $\pm 5\%$ or $\pm 10\%$ in each region.

CHAPTER 3.

Sound symbolism on synthetic speech continuum in judgment of size of visual stimulus

3.1 Introduction

Sound symbolism is the idea that sounds themselves form certain impressions (Ramachandran and Hubbard, 2001). For example, previous studies have shown that /p/, /s/, and /i/ have smaller impressions, whereas /b/, /o/, and /a/ have larger impressions (Newman, 1933; Sapir, 1929). In my previous study, I confirmed that /p/ and /i/ have smaller impressions and /b/ and /o/ have larger impressions, even when the sound stimulus is presented aurally (Itagaki and Kobayasi, 2017). Sound symbolism in nonspeech sounds has also been investigated, demonstrating that simple acoustic characteristics, such as frequency, contribute to sound symbolism (Gallace and Spence, 2006). Overall, although these studies suggested that both phoneme perception and simple acoustic characteristics are important, the relative contribution of each factor and the interactions between them have not been investigated systematically.

In this study, I sought to clarify the effect of the perception of a phoneme on sound symbolism using the stimulus continuum of speech sound. If the acoustic characteristic, rather than phoneme perception, is the main contributor to sound symbolism, the effect of sound symbolism should gradually change as the morphing rate changes. In contrast, if phoneme perception plays a measurable role in sound symbolism, the effects of sound symbolism should change in a stepwise manner around the middle of the stimulus

continuum, where perception switches from one phoneme to another. I also compare the effects of consonants and vowels and examine the differences in their effects.

3.2 Materials and Methods

3.2.1 Subjects

Ten subjects (6 females and 4 males; aged 21–24 years) participated in experiment 1, and 10 subjects (8 females and 2 males; aged 21–23 years) participated in experiment 2. All subjects were native Japanese speakers with no knowledge of either sound symbolism or the purpose of the experiments.

3.2.2 Visual stimuli

I examined the sound symbolism in terms of the magnitude of the visual stimulus (Gallace and Spence, 2006). Gray circles of different sizes were presented as the visual stimulus on a black background on a liquid crystal display (1024 × 768 lines, 15 inches; VL-150VA, Fujitsu) positioned 60 cm from the subject. There were three types of visual stimuli (one standard and two targets) and the target diameter was either smaller or larger than the standard by 40%. The standard comprised an outer circle of 300 pixels and an inner circle of 280 pixels. The larger target comprised an outer circle of 420 pixels and an inner circle of 400 pixels. The smaller target comprised an outer circle of 180 pixels and an inner circle of 160 pixels. The same visual stimuli were used in experiments 1 and 2. A red cross was always presented as the fixation point at the center of the screen.

3.2.3 Sound stimuli

The morphed speech sounds were presented as sound stimuli to examine the effect of phoneme perception on sound symbolism. The speech analysis and synthesis software “STRAIGHT” (Kawahara et al., 1999) was used to generate 10-step audio-morphed

continua between pairs of naturally recorded speech sounds. In experiment 1, I used a stimulus continuum between /pi/ and /ba/. In experiment 2, I used a stimulus continuum between /i/ and /a/. For instance, for each pair I generated nine intermediate stimuli in 10% acoustic steps from 0% (highly similar to /pi/ or /i/) to 100% (highly similar to /ba/ or /a/). Consequently, a total of 11 stimulus types (2 originals + 9 intermediates) were used in each experiment. Approximately 50% of the sounds were likely to be perceptually very ambiguous and could, for example, be randomly interpreted as /ba/ or /pi/.

3.2.4 Procedure

The subjects were required to estimate the size difference between the standard and target visual stimuli (larger or smaller). The same experimental paradigm was used for experiments 1 and 2. The trial began with a 2000 ms rest, then the standard stimulus was presented for 800 ms, and after an interstimulus interval (ISI) of 700 ms the target stimulus was presented for 500 ms, followed by a response period of 2000 ms (Fig. 3.1). During the rest and response periods, the screen was black. After the target stimulus was presented, the subject pressed the left or right button using his/her index or ring finger, respectively, to report the target size. Which button (left or right) represented which answer (larger or smaller) was alternated between subjects. I analyzed the RTs for each sound stimulus. There were 22 different combinations of overall stimulus types (two types of visual stimuli \times 11 types of sound stimulus). All the stimulus types were randomly ordered to form one block. One session contained a total of three blocks. In both experiments 1 and 2, each subject performed two sessions.

3.3 Results

I defined the congruent and incongruent conditions (described below) and analyzed the data accordingly. Under the congruent condition, the target stimulus was consistent with the impression of the concurrent sound (e.g., the larger target was presented with /ba/ (in experiment 1) or /a/ (in experiment 2)). In contrast, under the incongruent condition the target stimulus was inconsistent with the impression of the concurrent sound. I only used the data on the original sounds (i.e., 100% /ba/, 0% /ba/, 100% /a/, and 0% /a/) for the statistical analysis of the congruent and incongruent conditions.

3.3.1 Experiment 1: stimulus continuum /pi-/ba/

The mean Z score for RT under the incongruent condition was longer than that under the congruent condition, but the difference was not statistically significant (100% /ba/ vs 0% /ba/, $t = -1.26$; Fig. 3.2A). The differences in the Z scores for RT for all the sound stimuli are shown in Fig. 3.3A. In the stimulus in which the /ba/ component was less than 50%, RT for the smaller circle was shorter than that for the larger circle. In contrast, for the stimulus in which the /ba/ component was 50% or more, RT for the larger circle was shorter than that for the smaller circle. When the sound stimulus was 10% /ba/, RT for the larger circle was longer. This tendency may be an outlier or the value for 0% /ba/ might have been small. This should be considered in future research.

3.3.2 Experiment 2: stimulus continuum /i/–/a/

The mean Z score for RT under the incongruent condition was longer than that under the congruent condition, and the difference was significant (100% /a/ vs 0% /a/, $t = -2.48$, $p < 0.05$; Fig. 3.2B). The differences in the Z scores for RT for all the sound stimuli are shown in Fig. 3.3B. For the stimulus in which the /a/ component was less than 50%, RT for the smaller circle was shorter than that for the larger circle. In contrast, for the stimulus in which the /a/ component was 50% or more, RT for the larger circle was shorter than that for the smaller circle.

3.4 Discussion

In compare with the RT difference in Fig. 3.2A, the RT difference in Fig. 3.2B was large. This result suggested that the effect of vowel more affect to sound symbolism than the effect of consonant.

Comparing Figs. 3.3A and 3.3B, I see that the difference in response between the smaller circle and larger circle was more distinct in the vowel continuum (Fig. 3.3B) than in the combined consonant and vowel continuum (Fig. 3.3A). One possible explanation is that the effect of the consonant (i.e., /b/ or /p/) was only marginal, and the consonant masked the effect of the vowel (i.e., /a/ or /i/).

In Fig. 3.3B, the effect of sound symbolism (i.e., the difference between the closed symbol and the open symbol) did not change proportionally along the morph axis. Instead, the difference seems to fluctuate between 40% and 60%. This result suggests that the simple acoustic characteristic alone, which was manipulated as the morphing rate, was not the main contributor to the sound symbolism and that phoneme perception could play a measurable role in sound symbolism.

The data presented here have several shortcomings. First, when the behavioral response to the /pi/–/ba/ continuum was compared with that to the /i/–/a/ continuum, the results for the 50% morphing stimulus were inconsistent. One possible explanation for this inconsistency is that the perceptual boundaries of the phoneme were placed slightly differently between the two continua; i.e., in the /pi/–/ba/ continuum the boundary was between 40% and 50%, whereas in the /i/–/a/ continuum it was between 50% and 60%. This hypothesis can be tested by measuring the discrimination functions of these phoneme continua. In addition, the behavioral responses (Z scores) contain a point of discontinuity along the percentage axis. Specifically, the 10% morphing sound stimulus in the /pi/–/ba/

continuum produced a distinctly longer RT than the others (Fig. 3.3A). I cannot determine only from the current data whether the sample is an outlier. Therefore, further studies are required to understand the sound symbolism effect. Overall, this study indicates that my paradigm, combined with a morphed stimulus, provides a consistent behavioral response, and will assist in systematically evaluating the effect of phoneme perception on sound symbolism.

3.5 Summary

In many previous studies related to sound symbolism, stimuli were presented visually with alphabets, and subjects directly responded to questions regarding the impression of the sound. Therefore, it was not known whether basic acoustic features, phoneme perception, or both affect sound symbolism. In one study, it was suggested that factors such as length and frequency or the perception of the phoneme causes sound symbolism. The purpose of this study was to evaluate the effect of phoneme perception on the phenomenon. I created a phoneme continuum as the stimulus and focused on sound symbolism for visual size perception. Vocal sounds were assumed to convey the impression of a large or small size, according to previous research. Subjects were all right-handed native Japanese speakers, and they did not have any knowledge about the sound symbolism theory used in this experiment. The results indicated that when it was difficult to identify sound stimulus as a phoneme in experiments 1 and 2, sound symbolism was not observed; when the sound stimulus was clearly identified as a phoneme, sound symbolism was observed. These results suggest that phoneme perception is significantly related to sound symbolism.

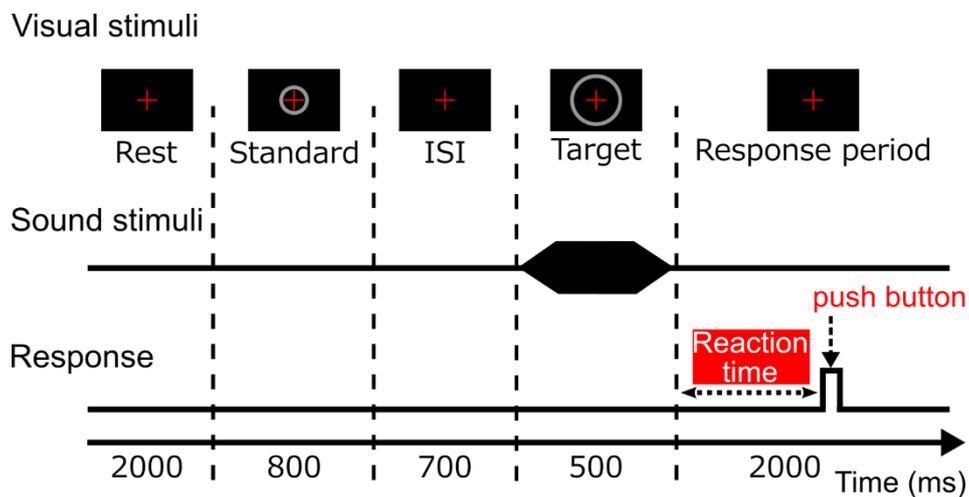


Figure 3.1 Illustration of the experimental design. Each trial started with a rest period (2,000 ms). The standard stimulus (800 ms) was then presented, followed by the target stimulus (500 ms) after a 700 ms interstimulus interval (ISI). Subjects were instructed to respond to the task during the response period (2,000 ms) via a button device, using the fingers of their left hand. A red cross was always presented as the fixation point.

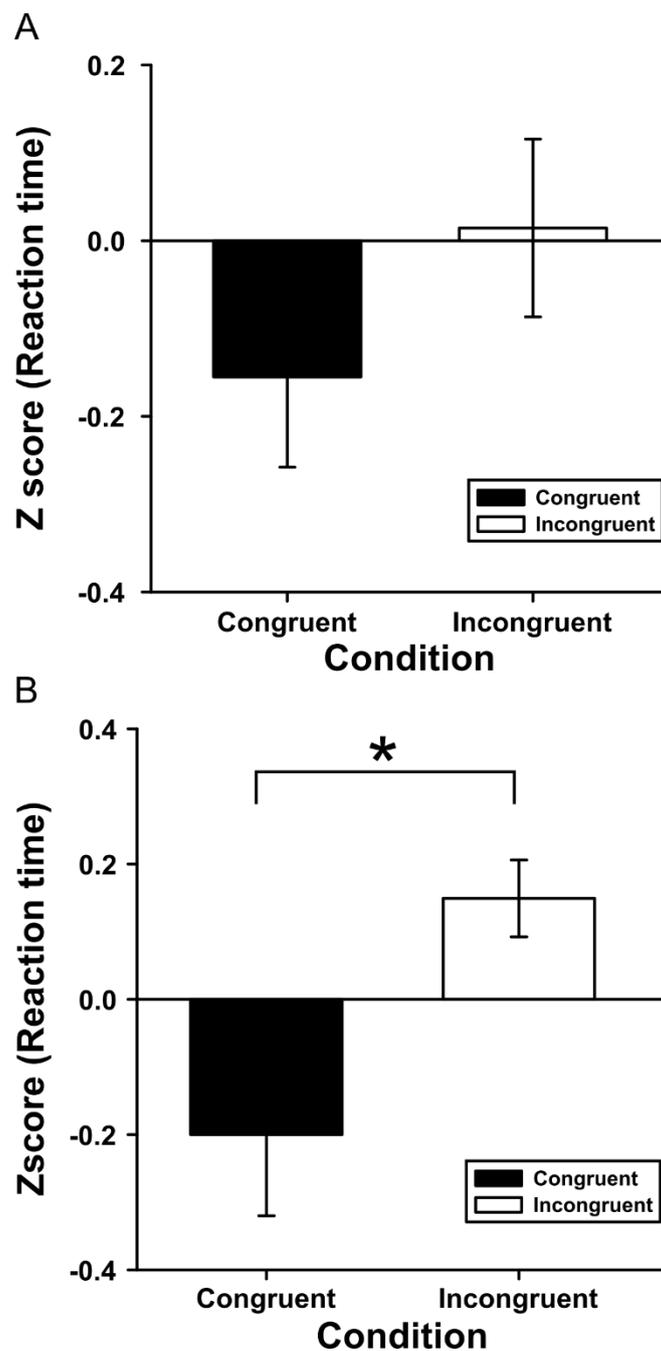


Figure 3.2 Average the RTs under each condition in both experiment 1 and 2 (Z-score). The vertical axis represents the Z-score and the horizontal axis represents the condition. Error bars represent the standard errors of the means. The mean Z score for RT under the incongruent condition was longer than that under the congruent condition (A) The RT difference when sound stimuli were 100% /ba/ and 100% /pi/ in experiment 1 ($t = -1.26$). (B) The RT difference when sound stimuli were 100% /a/ and 100% /i/ in experiment 2 ($t = -2.48, p < 0.05$).

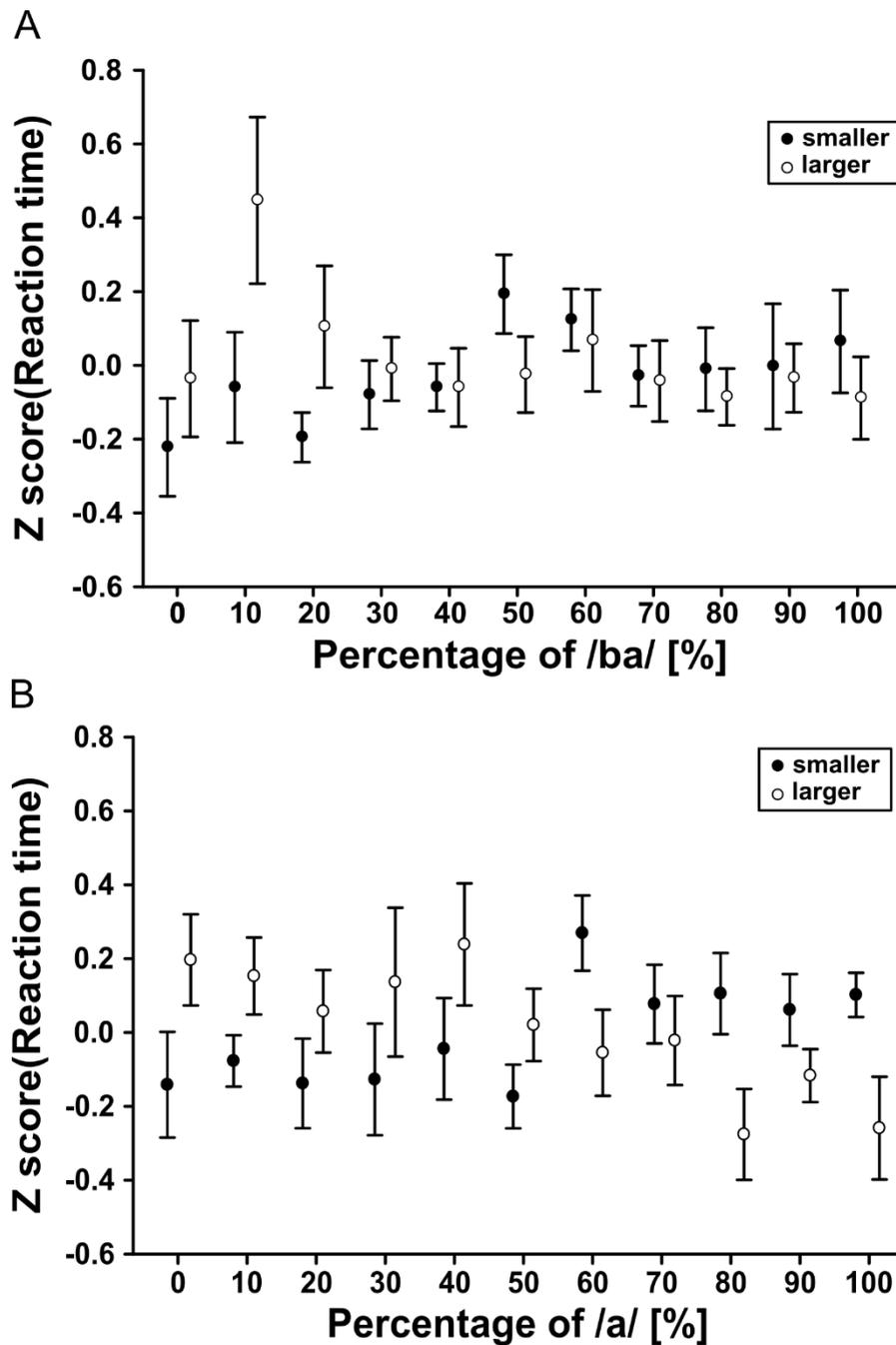


Figure 3.3 The differences in the Z scores for the RTs for all the sound stimuli. The vertical axis represents the Z-score and the horizontal axis represents the mixing percentage of sound stimuli. Error bars represent the standard errors of the means.

(A) Experiment 1: In the stimulus in which the /ba/ or /a/ component was less than 50%, RT for the smaller circle was shorter than that for the larger circle.

(B) Experiment 2: For the stimulus in which the /ba/ or /a/ component was 50% or more, RT for the larger circle was shorter than that for the smaller circle.

CHAPTER 4.

Brain regions involved in sound symbolism and their relationship with phoneme perception

4.1 Introduction

Sound symbolism has been confirmed as a form of word-meaning association by previous studies (Ramachandran and Hubbard, 2001; Sapir, 1929; Köhler, 1947). Sound symbolism is a phenomenon in which a sound conveys a particular impression. Several studies have confirmed the association between sounds and meanings to be “sound symbolism” or “phoneme symbolism.” For example, one study reported that a voiced consonant such as /b/ or the vowel /o/ convey the impression of largeness, unlike a voiceless consonant such as /p/ or the vowel /i/ that convey the impression of smallness (Newman, 1933). Other studies have reported that higher frequency sounds convey an impression of smallness, while lower frequency sounds convey the impression of largeness (Gallace and Spence, 2006; Ohtake and Haryu, 2013). There are many types of sound symbolism as shown by various studies (Fort et al., 2015; Noppeney et al., 2008; Westbury, 2005; Revill et al., 2014; Imai et al., 2015; Asano et al., 2015)

The key factors of sound symbolism include experience and articulatory movement. Knoeferle et al. (2017) indicated that when the higher first formant (F1) is paired with a longer vowel duration, it was associated with the impression of a greater size. Interestingly, the lower second formant (F2) was also associated with the impression of a greater size. In other words, the impression of largeness or smallness may be elicited to a

vowel's F2 frequencies. Furthermore, other studies confirm that articulatory movement also contributes to sound symbolism (Newman, 1933; Sapir, 1929; Shinohara and Kawahara, 2010). For example, vowels such as /a/, /u/, and /o/ convey the impression of larger-sized objects, whereas vowels such as /i/ or /e/ evoke the impression of smaller-sized objects. When we pronounce these vowels, the position of the tongue and the oral cavity differ. These differences may contribute to a vowel-size correspondence. On the other hand, when we pronounce the Japanese phoneme /a/, the position of the tongue is toward the back of the mouth and the oral cavity of the mouth is larger. On the other hand, when we pronounce the Japanese phoneme /i/, the position of the tongue is toward the front of the mouth, hits the upper jaw, and the oral cavity is smaller. An important theory to be considered while discussing these points is the motor theory, which suggests speech pronunciation is related to speech recognition (Liberman and Whalen, 2000). In one study, motor-evoked potentials were observed in relation to the anterior tongue muscles while participants were listening to words (Fadiga et al., 2002). In other words, the motor cortex related to the pronunciation of speech is also activated by passive listening to words. In my previous study, it was confirmed that brain regions related to phoneme perception contribute to sound symbolism (Itagaki et al., 2019). However, the study was unable to investigate the relationship between sound symbolism and phoneme perception directly. Therefore, the current study is focused on the relationship between speech (phoneme) perception and sound symbolism. As explained above, experience and articulatory movement are some of the factors of sound symbolism. However, these studies cannot investigate the individual effects of these factors, because the experimental paradigms used in such studies are complex. Research is needed to investigate these factors individually.

In my previous study, sound symbolism was observed even when the sound stimuli were phonemes. Speech sounds such as “bobo” and “pipi” were used as sound stimuli with simple syllables repeated twice. Therefore, this study was unable to investigate whether vowels or consonants have a greater effect on sound symbolism. Thus, this study investigated the difference between the effects of vowels and consonants. Furthermore, this study focused on the relationship between the motor cortex related to speech (phoneme) perception and sound symbolism. In this study, I also investigated how neural representation of visual stimuli is modulated by sound symbolism.

4.2 Materials and Methods

4.2.1 Subjects

Thirteen (one female and twelve males; aged 21–24 years) subjects participated in Experiment 1, while nine (all males; aged 21–24 years) of the initial 13 also participated in Experiment 2. All experiments were conducted with fMRI after obtaining written informed consent from all subjects. All subjects were right-handed and native Japanese speakers. None of the participants had any knowledge of sound symbolism or these experiments. In Experiment 1, data from two participants were excluded due to artifact or inadequate task performances (e.g., head moves > 3 mm).

4.2.2 Experimental apparatus

Each subject was positioned on an MRI scanner in a supine position; sound stimuli were presented via MRI-compatible headphones (Kiyohara Optics Inc., Tokyo, Japan), and visual stimuli were presented using the projector and mirror system, as in the previous study. Subjects viewed visual stimuli projected onto a mirror placed 18 cm away and in front of their eyes, and they pressed buttons (Current Designs, Inc., Pennsylvania, USA) with their left hand to respond. The control software (Presentation; Neurobehavioral System, Inc., Albany, CA, USA) for Experiments 1 and 2 were used to synchronize the experimental procedures with the fMRI scan.

4.2.3 Visual stimuli

4.2.3.1 Experiment 1

This experiment examined the effect of sound symbolism on judging the size of a visual stimulus, as in the previous study. The visual stimulus, a gray doughnut-shaped

circle , was presented on a frosted screen at the end of the scanner bore, which the subject could see via a mirror mounted on the head coil. The standard stimulus had an outer circle of 300 pixels and an inner circle of 280 pixels. The target stimulus was either smaller or larger than the standard one by $\pm 40\%$ of its diameter. In total, three sizes (one for the standard and two for the targets) were used. The standard stimulus was presented for 250 ms, and the target stimulus was presented for 700 ms. A red cross (34 pixels) was always presented as a fixation point at the center of the screen.

4.2.3.2 Experiment 2

This experiment's task was the phoneme discrimination. In this experiment, the instruction screen was presented as the visual stimulus and "please choose the sound heard" was presented as the instruction phrase. These instruction screens were presented for 4500 ms, and a red cross (34 pixels) was presented as a fixation point at the center of the screen at other times.

4.2.4 Sound stimuli

The sound stimuli were click sounds, white noise, and phonemes (/bo/, /bi/, /po/, and /pi/). A publicly available sound dataset (FW03; NTT communication Science Laboratories, Kanagawa, Japan) was used for these phonemes. All phoneme sounds were recorded at a sampling frequency of 48 kHz and a quantization of 16 bits. The single-syllable utterances /bo/, /bi/, /po/, and /pi/ were spoken by a male. The duration of sound was 250 ms, and the stimulus amplitude was 64 dB sound pressure level (SPL). The hearing level of /bi/ was louder than /po/ by approximately 8.14 dB (Nagatani et al., 2008). The sound stimulus was synchronized with the visual stimulus. The click sound was

always presented with sound 1, and one of the sound stimuli was presented randomly with sound 2. In my previous study, I confirmed that /bo/ conveys a greater impression of largeness than /pi/.

4.2.5 fMRI parameters

These experiments used the same fMRI parameters used in my previous study. Functional images of brain activity were acquired using a 1.5-T MRI system (Echelon Vega, Hitachi Medical Corporation, Tokyo, Japan) as T2*-weighted images using a gradient echo-echo planar imaging (GE-EPI) sequence with a resolution of $3 \times 3 \times 5$ -mm voxels (30 slices; FOV: 192 mm; matrix: 64×64 ; TR: 3000 ms; TE: 50 ms; FA: 90 degrees). Five initial scans were discarded to avoid magnetic saturation effects. A structural T1 image was scanned at a $1 \text{ mm} \times 1 \text{ mm} \times 1\text{-mm}$ resolution using a three-dimensional gradient echo inversion recovery (3D-GEIR) sequence (192 x 1-mm thick slices; sagittal; FOV: 256 mm, matrix: 256×256 ; TR: 9.7 ms; TE: 4 ms; T1: 1045 ms; FA: 8 degrees).

4.2.6 Procedure

4.2.6.1 Experiment 1

Subjects were asked to judge the difference in sizes of standard and target stimuli, as in the previous study. Each trial began with a 3,000-ms rest period. Then, the standard stimulus was presented for 250 ms, followed by the ISI for 300 ms. Next, the target stimulus was presented for 700 ms, followed by a response period of 800 ms (Fig. 4.1A). Subjects were required to answer within the time that the target stimulus was presented. The screen was black during the rest and response periods. When the subjects responded

to the task by pressing a button, the next trial began automatically. In the control task, subjects were instructed to press the middle button with their middle finger, regardless of the visual stimulus, when the sound (sound 2) presented with the target stimulus was identical to the sound (sound 1) presented with the standard stimulus. In the comparison task, subjects were directed to respond (using their index and ring fingers to press the left and right buttons, respectively) according to the size of the target circle compared with the standard size (Fig. 4.1B) when the sound stimulus presented with the target differed from that presented with the standard stimulus. The button assignment (i.e., which button represented which answer) was changed between subjects. There were two combinations of visual stimuli (one standard \times two targets) and six combinations of sound stimuli (one Sound 1 \times six Sound 2), yielding a total of 12 stimulus combinations. In Experiment 1, each subject completed 384 trials (control task = 144 trials; comparison task = 240 trials). Each subject conducted one training session.

4.2.6.2 Experiment 2

In this experiment, subjects were required to discriminate between the heard sounds. Each trial began with a 1,500 ms rest period. Then, the instruction screen was presented for 4,500 ms. The sound stimulus was presented 1,000 ms after the instruction screen was presented, followed by a response period of 1,500 ms (Fig. 4.1C). The screen was black during the rest period.

4.2.7 Analysis

Functional imaging data were processed using the SPM12 software (Wellcome Department of Cognitive Neurology, London, UK). For realignment, the image was

spatially normalized to an EPI template in the Montreal Neurological Institute (MNI) space and smoothed with an 8-mm full width at half maximum Gaussian kernel. To examine the brain activation related to sound symbolism and speech perception, I created different types of contrast images (congruent minus incongruent, incongruent minus congruent, speech perception and discrimination). These effects were modeled as stick functions with a convolution of the canonical hemodynamic response function. Movement parameters from realignment corrections were included in the model as covariates of no interest. Significant regions were superimposed on a standard human brain template from MRIcro software (www.mccauslandcenter.sc.edu/crnl/mricro). Furthermore, FSL software version 5.0 (FMRIB Software Library, FMRIB, Oxford, UK; Smith et al., 2004) was used for physiological noise removal before analysis (Aso et al., 2017). Region-of-interest (ROI) analysis was performed using the SPM toolbox.

4.2.8 Physiological noise removal

A denoising approach was applied to minimize pseudo-positives and inter-individual variation of the fMRI due to physiological noise (Aso et al., 2017; Erdoğan et al., 2016; Tong et al., 2013). This technique used to reduce noise involves tracking the regional phase variation of the low-frequency oscillation of systemic origin.

First, the oscillation time course is captured by selecting a set of voxels that give a cross-correlogram peak at precisely zero using a bandpass filtered (0.08–0.07 Hz) global signal.

Based on the averaged time-series across these voxels (with lag = 0) as the new reference, the next set of voxels was located by finding those with a cross-correlogram peak at 0.5 s (0–0.5 s) indicating a perfusion time lag. This recursive procedure was

repeated up to 7 s both upstream and downstream to create a lag map and the corresponding time-series set. This spatiotemporal lag structure was subsequently regressed out from the original data, which can be considered as global signal regression tailored for each voxel but affecting only the slow components.

4.3 Results

Eleven subjects performed well on the comparison task; the mean correct response rate was 95.1%. I analyzed the RTs for the comparison tasks. As my previous study and some other previous studies have confirmed that /b/ and /o/ elicited an impression of largeness and /p/ and /i/ elicited an impression of smallness, I assumed that /bo/ and /pi/ would induce impressions of largeness and smallness, respectively. I defined the congruent and incongruent conditions as follows and analyzed the data accordingly. In the congruent condition, the target stimulus was consistent with the reaction to the sound (i.e., the larger target was presented with /bo/ or the smaller target with /pi/). In the incongruent condition, the target stimulus was inconsistent with the reaction to the sound (i.e., the larger target was presented with /pi/ or the smaller one with /bo/).

First, I examined the influences of consonants and vowels and examined the different effects they had on sound symbolism. Upon examining the influence of consonants, I determined that /bo/ and /bi/ convey an impression of largeness, while /po/ and /pi/ convey an impression of smallness. Regarding the influence of vowels, I found that /bo/ and /po/ convey an impression of largeness, while /bi/ and /pi/ convey an impression of smallness. In both cases, the mean RT in the congruent condition was shorter than that in the incongruent condition. However, the influence of vowels was larger than that of consonants (Fig. 4.2). Regarding the influence of vowels, the difference between the mean RT in the congruent condition and incongruent condition was significant ($t = -3.15$, $p < 0.05$, paired t-test with Bonferroni correction).

In this study, I examined the brain activity related to sound symbolism and speech perception by contrasting the incongruent condition with the congruent condition and conducting an ROI analysis. The left postcentral gyrus was more activated in the

incongruent condition than in the congruent condition (Fig. 4.3). The anterior cingulate cortex (ACC), which was confirmed to be associated with sound symbolism in my previous study, was more activated in the incongruent condition than in the congruent condition. However, the difference in activation intensity was not significant.

I also investigated the brain region related to motor control during speech perception using ROI analysis. First, the peak value of the contrast estimate under f-contrast in Experiment 2 was obtained after masking the left/right precentral and postcentral gyri using Neuromorphometrics masks created using the SPM toolbox (Table 4.1). Next, I created ROIs from these peak coordinates. The ROIs were spheres with a radius of 8 mm. Then, to examine the relationship between sound symbolism and speech perception, contrast estimates in Experiment 1 were masked with these ROIs and plotted (Fig. 4.4). The region related to the left postcentral gyrus observed in Experiment 1 demonstrated activation during speech discrimination, and this effect of interaction between vowel and visual size was significant [$F(1, 10) = 6.328, p = 0.031$].

I also investigated whether sound symbolism influences the visual cortex. The activity of the visual cortex in symbolic sound experiments was measured by masking the ROIs created using the peak coordinate of the visual cortex obtained from half of the trials to the contrast estimate of the other half of the trials. First, the peak coordinate of the contrast estimates from the visual cortex in each subject in Experiment 1 was measured using Neuromorphometrics masks created by the SPM toolbox. Then, visual cortex activity was measured using ROIs created by the above coordinate in each subject and plotted (Fig. 4.5). In V1 to V3, the values of the contrast estimates in the congruent condition were larger than that in the incongruent condition and greater in the larger circle condition than in the smaller circle condition.

4.4 Discussion

The behavioral results suggest that vowels affect sound symbolism more strongly than do consonants. In Japanese, consonants are always paired with a vowel. Therefore, vowels may be more important than consonants when we pronounce words. Furthermore, this result suggests that sound symbolism was observed even when sound stimuli consisted of a single syllable.

Brain activity in the left postcentral gyrus was observed in Experiments 1 and 2 both. Previous studies indicate that this region is related to tongue movement and speech (Ehrsson et al., 2003; Grabski et al., 2012). Ehrsson and his colleagues (2003) investigated whether imagining voluntary movements of different body parts activates corresponding motor representations. In the movement condition of this study, subjects were asked to make horizontal movements with their tongue and imagine its movement in the imagination condition. The left and right precentral and postcentral gyri were activated in both the movement and imagination conditions. These regions were very close to the regions observed in Experiment 2 of the current study. Grabski and her colleagues (2012) examined brain regions related to the lips, jaw, larynx and tongue movement. In this study, subjects were required to pronounce a vowel and move each body part (lip, jaw, larynx, and tongue). The left premotor cortex was activated in the vowel voicing and tongue movement conditions. This region was also very closely located to the regions observed in Experiment 2 of the current study. Thus, Grabski and her colleagues (2012) suggested that the premotor cortex and the precentral and postcentral gyri contribute to speech perception and discrimination. Again, this region was also observed in Experiment 1 of my study, which investigated sound symbolism. This finding suggests that the motor areas related to speech affect sound symbolism.

Moreover, this region was more activated in the incongruent condition than in the congruent condition in Experiment 1, suggesting that the incongruent condition was more difficult for perception and speech discrimination. In other words, the phonetic symbol may have boosted speech recognition and visual size determination.

The ACC was more activated in the incongruent condition than in the congruent condition, similar to the results of my previous study. However, the difference in the intensity of activation was not significant. In this study, the sound stimulus was a single syllable with linguistic reduction unlike that in my previous study which used speech with a single syllable repeated twice. In addition, the experimental paradigm of this study was more complex than that of my previous study. Therefore, the difference in ACC activity in this study is considered to be smaller than in my previous study.

In V1 to V3, the values of the contrast estimate in the congruent condition were larger than that in the incongruent condition. Additionally, the values of the larger circle were greater than that of the smaller circle in both congruent and incongruent conditions. Previous studies have considered size perception to be a high-level function, probably mediated by specific cortical areas selective to objects and object size (Ungerleider and Bell, 2011; see Kourtzi and Connor, 2011, for a review; Konkle and Oliva, 2012). Furthermore, the different primary visual cortex representations were activated by the apparent size of an object when the object was presented together with a perspective visual cue (Murray et al., 2006; Fang et al., 2008; Pooresmaeili et al., 2013; Schwarzkopf and Rees, 2013). In other words, different parts of the visual cortex were activated by subjective perception. Meanwhile, the visual cortex was also activated by listening to mimic words (Osaka, 2009). In the walk condition of this study, subjects were asked to listen to walk-evoking mimic words and nonsense syllables in the control

condition, and to close their eyes while listening. The visual cortex was more activated in the walk condition than in the control condition. This result suggests that a visual image was recalled from the mimic words. In my study, the values of the contrast estimate in the congruent condition were larger than that in the incongruent condition. These results are consistent with the results of previous studies. In addition, the difference between congruent and incongruent conditions was observed in the lower order area of the visual cortex. It can be inferred from this result that the first step of speech symbolism may be processed in the lower levels of the visual cortex, such as V1 to V3.

These results confirm that acoustic character is not the only factor involved in sound symbolism. Furthermore, it suggests that the motor area related to speech perception also affects sound symbolism. However, the relationships among sound symbolism, speech processing, and acoustic character should be investigated by further research.

4.5 Summary

In this study, two experiments were conducted to investigate the relationship between speech perception and sound symbolism. In Experiment 1, I focused on brain activity related to sound symbolism for visual size. Subjects were required to identify the visual difference in size between a standard and a target stimulus, while listening to sound stimuli. Sound stimuli consisted of a single syllable (/bo/, /bi/, /po/, and /pi/), a noise, and a click sound. In Experiment 2, subjects were asked to discriminate between heard sounds. Only the Japanese phonemes used in Experiment 1 were used as sound stimuli in Experiment 2. The behavioral data showed that sound symbolism was observed even when the sound stimulus was a single syllable. Furthermore, when comparing the RT difference between incongruent and congruent conditions, vowels were found to have had a greater influence than consonants. The fMRI data showed that the left postcentral gyrus was more activated in Experiments 1 and 2 both. In other words, the brain regions related to speech perception also contributed to sound symbolism. In addition, the values of the contrast estimate in the congruent condition were larger than that in the incongruent condition in the lower order of the visual cortex. From this result, it can be inferred that the first step of speech symbolism may be processed in lower levels of the visual cortex, such as V1 to V3. These results confirm that acoustic characters are not the only factors involved in sound symbolism, suggesting that the motor area related to speech perception also influences sound symbolism.

Brain region	MNI coordinates			Z-score	Cluster size (voxels)
	x	y	z		
① Left precentral gyrus	-40	-30	64	4.26	911
② Left postcentral gyrus	-62	-2	28	2.44	65
③ Left precentral gyrus	-50	4	40	2.16	44
④ Right precentral gyrus	36	-24	66	4.07	591
⑤ Right postcentral gyrus	56	0	34	2.52	225

Table 4.1 MRI activation clusters related to motor area derived speech perception. Note: Peaks that survived the threshold at voxel level $p < 0.1$, uncorrected, and cluster extent $k > 40$ voxels uncorrected. This table shows the MRI activation cluster related to motor area derived speech perception.

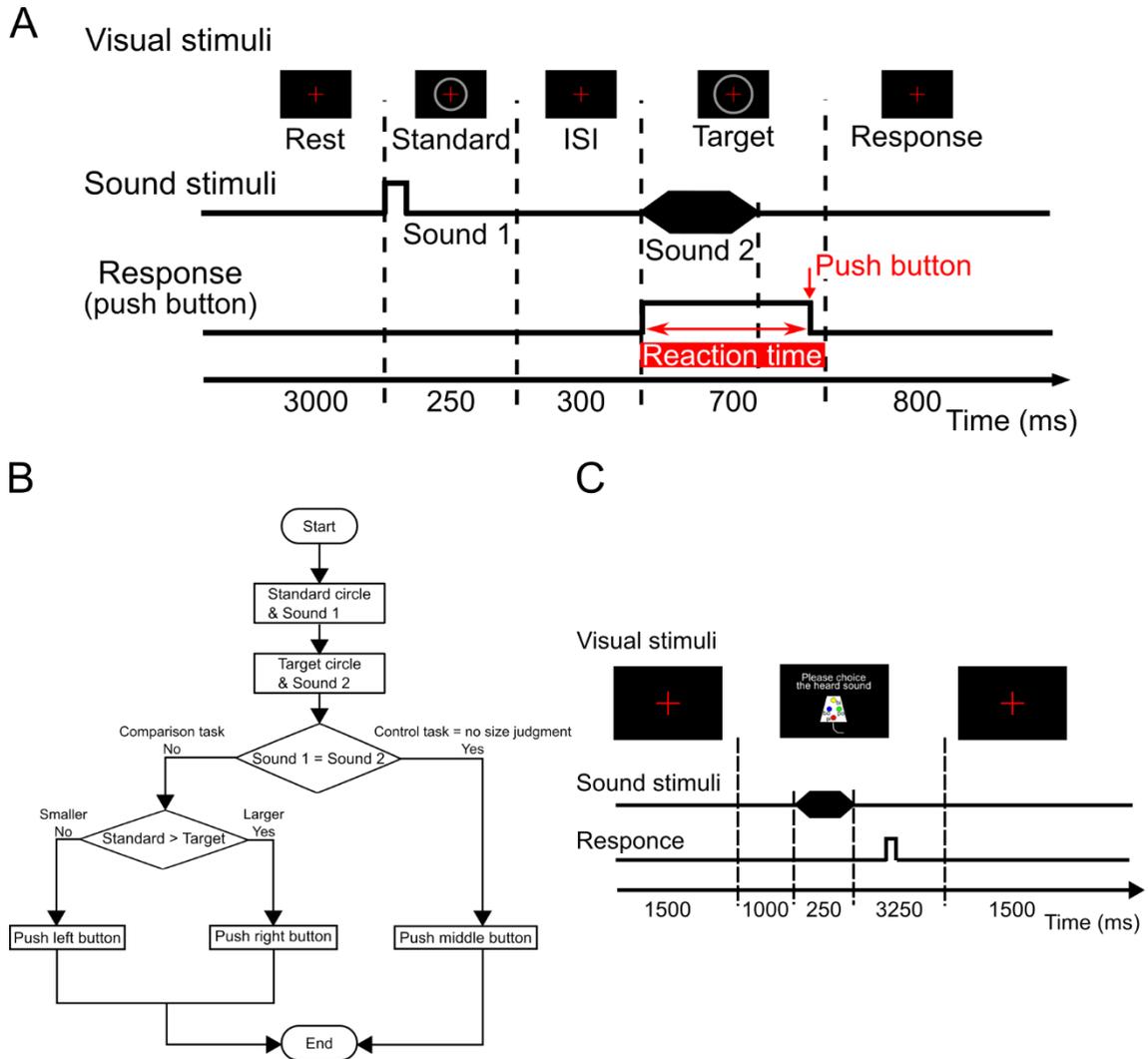


Figure 4.1 Illustration and flow chart of the experimental design in Experiment 1 and 2. **(A)** Illustration of the experimental design in Experiment 1. Each trial started with a rest period (3,000 ms). The standard stimulus (250 ms) was then presented, followed by the target stimulus (700 ms) after a 300 ms interstimulus interval (ISI). Subjects were instructed to respond to the task during the response period (800 ms) via a button, using the fingers of their left hand. A red cross was always presented as the fixation point. **(B)** Flow chart illustrating the experimental process. Subjects were instructed to push the middle button to respond to the control task and the left or right button to respond to the comparison task. Button assignment (which button represented a specific answer) in the comparison task was changed between subjects. **(C)** Illustration of the experimental design in Experiment 2. Each trial started with a rest period (1,500 ms). Then, the instruction screen was presented (4,500ms). The sound stimulus was presented 1,000 ms after the instruction screen was presented and was followed by a response period (1,500 ms). A red cross was presented as the fixation point.

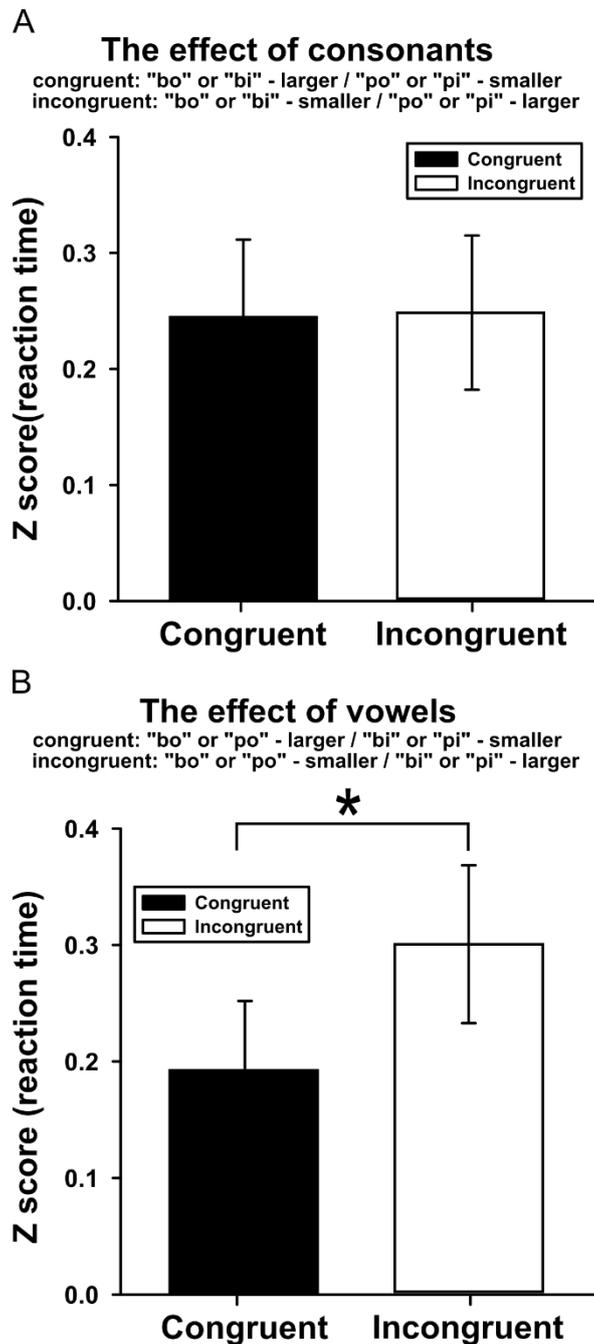


Figure 4.2 Average the RTs for each condition (Z-score). The vertical axis represents the Z-score and the horizontal axis represents the condition. Error bars represent the standard errors of the means. The Z-score was greater in the incongruent condition than in the congruent condition in both conditions.

(A) This figure shows the influence of consonants ($t = -0.073$).

(B) This figure shows the effect of vowels.

The difference in the vowel effect condition was significant ($t = -3.15, p < 0.05$, paired t-test with Bonferroni correction.)

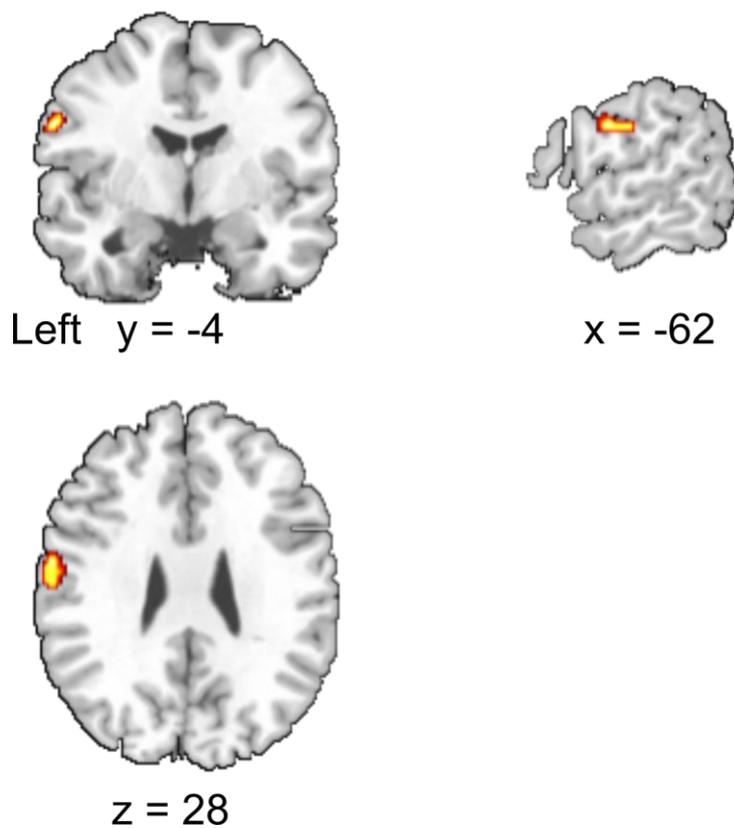


Figure 4.3 Activation maps in the incongruent condition. The left postcentral gyrus exhibited greater activity in the incongruent condition than in the congruent condition (threshold at voxel level $p < 0.01$, uncorrected, and cluster extent $k > 50$ voxels uncorrected).

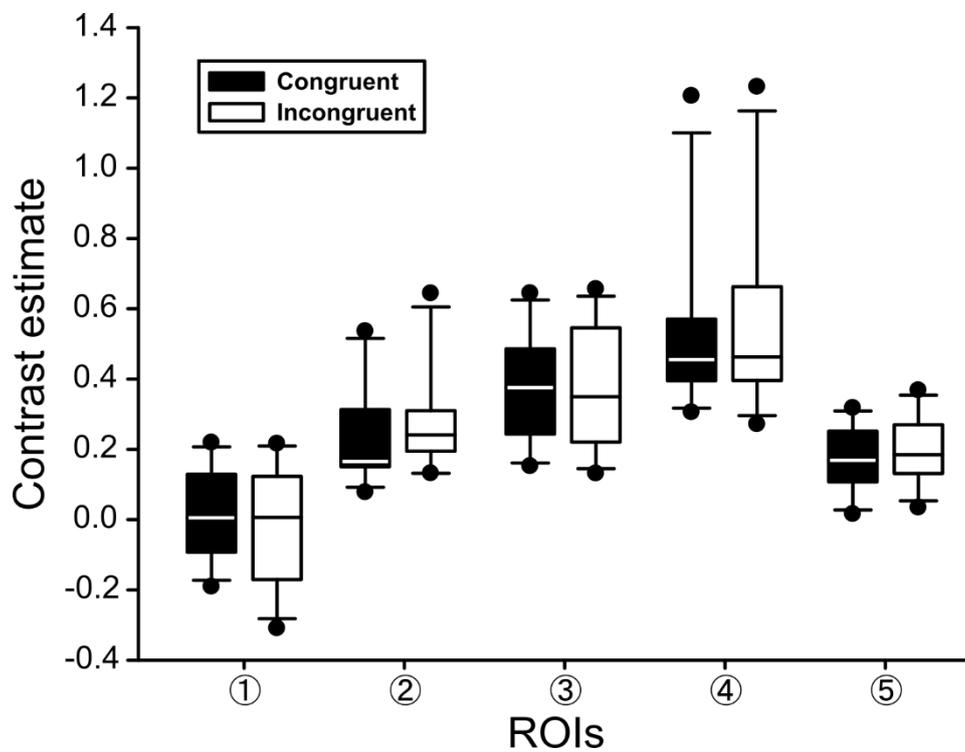


Figure 4.4 Contrast estimate for each condition in each ROI related to the left/right precentral and postcentral gyri. The vertical axis represents the contrast estimate and the horizontal axis represents each ROI. The ROIs were spheres with a radius of 8 mm. Error bars represent the standard errors of the means. Each number on the horizontal axis represents each region in Table 4.1. In the left postcentral gyrus, the value of the incongruent condition was greater than that of the congruent condition (②).

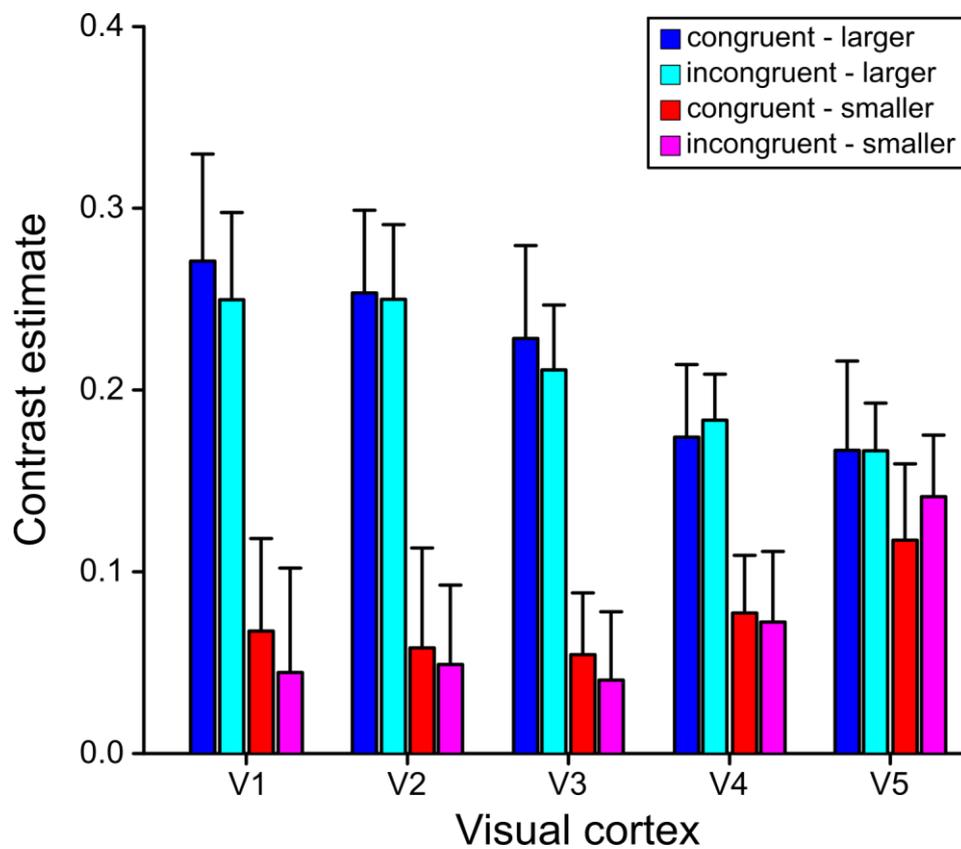


Figure 4.5 Contrast estimate for each condition in the visual cortex during Experiment 1. The vertical axis represents the contrast estimate and the horizontal axis represents each area of the visual cortex. Error bars represent the standard errors of the means. In V1 to V3, the values of contrast estimates in the congruent condition were larger than that in the incongruent condition. Values were also greater in the larger circle condition than in the smaller circle condition.

CHAPTER 5.

Change of sound symbolism during the language development: Infants' preferential looking behavior

5.1 Introduction

There are several phases in the process of language learning (Kuhl, 2004). Infants learn to make sounds using the mouth and subsequently learn the relationship between words and their meanings. More complex communication is later possible by learning the grammatical structure. As mentioned above, infants increase their vocabulary by learning the connection between sounds or letters and their meanings or notions. Such word-to-meaning mapping has been considered arbitrary, however, phenomena such as sound symbolism appear to contribute to word-to-meaning mapping (de Saussure, 1983; Sapir, 1929; Köhler, 1947; Ramachandran and Hubbard, 2001). Sound symbolism refers to a phenomenon in which the sound itself conveys a specific impression, for example, a high frequency sound conveys an impression of smallness, while a low frequency sound conveys an impression of largeness (Gallace and Spence, 2006; Ohtake and Haryu, 2013; Hirata et al., 2011; Shinohara et al., 2016; Fort et al., 2015; Knoeferle et al., 2017).

Various studies have examined the effect of sound symbolism on experience and learning (Lockwood et al., 2016; Imai et al., 2008; Nygaard et al., 2009; Kantartzis et al., 2011; Imai et al., 2015). One study investigated whether subjects could infer the meaning of an ideophone from an unfamiliar language (Lockwood et al., 2016). Specifically, after the Japanese ideophone was presented, the translation opposite to the real translation of

the ideogram was presented in Dutch. Dutch subjects were required to select the real translation of their ideophones. According to the results, subjects selected real translations for unknown ideophones with a degree of accuracy above the level of chance. Another study examined whether onomatopoeia created for 2- and 3-year-olds was associated with the manner of walking expressed by the words (Imai et al., 2008). Specifically, the children were asked to select a video that matched the words they heard from two videos displayed side-by-side showing different manners of walking. The videos that matched the words were selected at a rate above the level of chance. These results suggest that sound symbolism contributes to the determination of the meaning of unfamiliar words.

Sound symbolism has also been observed in 4-month-olds (Peña et al., 2011). One study examined the connections between phonemes (/i/, /e/, /o/, and /a/) and different sizes of geometric shapes (circle, oval, triangle, and square) in Spanish infants. The infants tended to look longer at small shapes when the sound stimuli were /i/ or /e/ compared to when the sound stimuli were /o/ or /a/. Meanwhile, infants tended to look at large shapes longer when the sound stimuli were /o/ or /a/ compared to when the sound stimuli were /i/ or /e/. This suggests that an association between phonemes and visual size exists before speech production in 4-month-olds. That is, sound symbolism may be acquired before speech production, and if it is acquired prior to speech production, it could be considered to contribute to language development. In addition, another study investigated the relationship between the frequency of sounds and visual size in 4- and 6-month-olds (Fernández-Prieto et al. 2015). In this study, infants were presented with a frequency-modulated sound (from 300 Hz to 1,700 Hz and from 1,700 Hz to 300 Hz) and a circle which changed size in tandem with the frequency modulation. Six-month-olds tended to look at the visual stimuli longer when they matched the impression conveyed

by the sound while no difference in looking time across conditions was observed in 4-month-olds. These and other studies have provided evidence to support that sound symbolism occurs before speech production in infants (Imai et al., 2015; Maurer et al., 2006; Ozturk et al., 2013; Asano et al., 2015). However, since these experiments used different experimental paradigms, they were not able to comprehensively examine the relationship between sound symbolism and the language development process. Therefore, experiments should be conducted across various ages to explore whether sound symbolism changes with language development.

To investigate the relationship between language development and sound symbolism in this study, I used the same paradigm for infants of various ages below and above 12 months, which is typically considered to be the speech production period when many infants produce their first word. In order to evaluate this using the same paradigm, the looking time for each stimulus was measured using an eye tracker. It was asserted that the same paradigm can be used for various ages by using the infant's preferential looking method to examine whether sound symbolism changes with language development. If sound symbolism changes as infants acquire language, then sound symbolism may contribute to language development itself.

5.2 Materials and Methods

5.2.1 Subjects

Thirty-eight 9-month-old infants (range, 8 months 24 days to 9 months 7 days; 17 boys, 21 girls), thirty-three 12-month-olds (range, 11 months 24 days to 12 months 7 days; 15 boys, 18 girls), and thirty-one 15-month-olds (range, 14 months 24 days to 15 months 7 days; 16 boys, 15 girls) participated in the study. All parents of infants who were participants gave written informed consent prior to participation. They were native speakers of Japanese. The Japanese MacArthur-Bates Communicative Development Inventory was used to evaluate the language development stage at each age (Ogura et al. 2004). The average number of understood words in the vocabulary section of the questionnaire was 31.9 words (SE = 5.8) in 9-month-old infants, 66.0 words (SE = 8.7) in 12-month-olds, and 102.3 words (SE = 12.4) in 15-month-olds.

5.2.2 Experimental apparatus

In this study, infants' behavior was observed while sitting on their parents' lap, positioned 65 cm from the display. The eye movement of infants while they watched the video was measured using an eye tracker (Tobii, Tobii pro X3-120). Visual stimuli were presented on the display (DELL, 1800FP), and at the same time, sound stimuli were presented from the speaker behind the display (BOSE, Computer Music Monitor).

5.2.3 Visual stimuli and sound stimuli

5.2.3.1 Experiment 1

Visual stimuli

Experiment 1 included two conditions: a shape condition and a size condition. Two

shapes with different outlines (rounded and spiky shapes) were used as visual stimuli in the shape condition, and pictures of balls with different sizes (large and small balls) were used as visual stimuli in the size condition. In the shape condition, the visual stimulus flashed in tandem with the sound stimulus, and, in the size condition, the visual stimulus rotated in tandem with the sound stimulus. Two types of figures or pictures were presented side-by-side. In order to account for infants' the left or right screen preference, the location of the visual stimulus (left or right side) was counterbalanced within participants. There were 16 combinations in total (four visual stimuli x four sound stimuli). In Experiment 1, the length of one video was 12 seconds.

Sound stimuli

In Experiment 1, the sound stimuli consisted of low frequency (100 Hz) and high frequency (400 Hz) tone bursts with a harmonic structure and the Japanese phonemes /a/ and /i/ spoken by two native Japanese females. The amplitudes of the sounds were adjusted to 74 dB sound pressure level (SPL) (± 1 dB SPL). The length of the sound stimulus was approximately 250 ms.

5.2.3.2 Experiment 2

Visual stimuli

Two faces, one with rounded eyes and one with spiky eyes, were used as visual stimuli in experiment 2. The eyes of the face flashed in tandem with the sound stimulus. The length of one video was 9 s, and, in Experiment 2, there were a total of eight combinations (two visual stimuli x four sound stimuli). Experiment 2 included two types of faces presented side-by-side, and the position of the visual stimulus (left or right) was

counterbalanced within participants, as in Experiment 1.

Sound stimuli

In Experiment 2, four unfamiliar four-mora words (“kinkiri,” “chingai,” “moudan,” and “yumon”) spoken by two native Japanese females were used as sound stimuli. The results of an evaluation completed by adult participants indicated that these four words were selected as the top two words associated with the impression of roundness and the top two words associated with the impression of spikiness. In this impression evaluation experiment, eight words (“chingai,” “mongara,” “sekiten,” “moudan,” “hannaga,” “kinkiri,” “chinrin,” and “yumon”) from the voice data set (FW03; NTT communication Science Laboratories, Kanagawa, Japan) were used as sound stimuli. Using the two-alternative forced choice method, subjects were required to answer which of the two words randomly extracted from these eight words best matched the spiky shape. Differences in the order of playback constituted different combinations, and each combination was presented three times. If a word matched a spiky shape, the selected word was scored as one point and the total score of each word was calculated. The sharpness of each word was determined by the scores. Using this method, four words (“chingai,” “kinkiri,” “moudan,” and “yumon”) were selected as sound stimuli for Experiment 2. These sound stimuli were also adjusted to 74 dB SPL, and the length of the sound stimuli was 750 to 870 ms.

5.2.4 Procedure

The experimental environment was the same in Experiment 1 and Experiment 2. Infants were required to sit on their mother’s lap, positioned approximately 65 cm from a

display, and were required to look at the display. The movement of the infants' eyes at this time was measured by the eye tracker positioned at the bottom of the display. In order to prevent the mother's behavior from influencing the infant's behavior, the mother was required to cast her eyes down and not look at the visual stimulus. The sound stimulus was presented simultaneously with the visual stimulus. In order to acquire the attention of infants, videos that did not affect the experiment were presented before and after the experiment and between each stimulus (Fig. 5.1). Experiment 1 was divided into two sessions (eight videos in each session), and three sessions were conducted in combination with Experiment 2. All videos were presented randomly at each session, and each session was also presented randomly across subjects.

5.2.5 Analysis

Subjects who cried during the experiment, looked at only one of the stimuli presented side-by-side, or had an insufficient gaze rate during the experiment (30% or less) were excluded from the following analysis. Therefore, seventeen 9-month-olds, eight 12-month-olds, and five 15-month-olds were excluded from the shape condition, and fifteen 9-month-olds, seven 12-month-olds, and two 15-month-olds were excluded from the size condition in Experiment 1. In Experiment 2, thirteen 9-month-olds, six 12-month-olds, and seven 15-month-olds were excluded from subsequent analyses. In this study, the duration of time that infants spent looking at the visual stimuli was analyzed. First, areas of interests (AOIs) large enough to include the visual stimulus were set equal to each visual stimulus. Then, if a line of sight entered the AOIs, it was considered that the visual stimulus was being watched, and the total looking time spent in the AOIs was measured.

5.3 Results

In this experiment, the looking time for each stimulus was analyzed. Previous research has indicated that non-sense words including /a/, the phoneme /a/, and low-frequency sounds were associated with impressions of roundness and largeness, whereas, non-sense words including /i/, the phoneme /i/, and high-frequency sounds were associated with impressions of spikiness and smallness (Newman, 1933; Auracher et al., 2009; Gallace and Spence, 2006; Nielsen and Rendall, 2013; Lockwood and Dingemans, 2015). I defined the congruent and incongruent conditions as follows and analyzed the data accordingly. In the congruent condition, the visual stimulus was consistent with the impression of the sound (i.e., the larger ball was presented with /a/ or the smaller ball with /i/). In the incongruent condition, the visual stimulus was inconsistent with the impression of the sound (i.e., the rounded shape was presented with /i/ or the smaller one with /a/).

In Experiment 1, the difference in the looking time between visual stimuli for each sound stimulus (rounded shape - spiky shape in the shape condition, large ball picture - small ball picture in the size condition) was compared (Fig. 5.2). When the sound stimulus was a tone burst, the amount of time spent looking in the incongruent condition was found to be longer than that in the congruent condition in both the shape and size conditions, regardless of the age of subjects. However, when the sound stimulus was a Japanese phoneme, the amount of time 9-month-olds spent looking in the congruent condition was longer than the incongruent condition. In contrast, 12-month and 15-month-olds tended to look longer in the incongruent condition than in the congruent condition. This tendency was observed for both the shape and size conditions. When comparing the looking time for each stimulus at each age of the infants, the spiky shape tended to be more preferred than the rounded shape in the shape condition regardless of the age of subjects, and the

large ball was also more preferred than the small ball in the size condition (Figs. 5.3 and 5.4). The statistical significance was not observed in these results.

In Experiment 2, I compared the difference in looking time between visual stimuli for each word (face with rounded eyes-face with spiky eyes; Fig. 5.5). The impression evaluation completed by adults indicated that “kinkiri” was the word that conveyed the strongest impression of pointiness, followed by “chingai,” “moudan,” and “yumon”. In this study, I determined that the two words “kinkiri” and “chingai” give the impression of spikiness, while the two words “moudan” and “yumon” give the impression of roundness. The results show that when the word was “chingai” the time spent looking at the face with rounded eyes was the longest, regardless of the age of subjects. When the word was “kinkiri,” subjects tended to spend more time looking at the face with spiky eyes, regardless of their age. Unlike Experiment 1, there was no difference in the looking time between the congruent condition and the incongruent condition. However, a consistent preference between the word and time spent looking at the face was seen regardless of the age of infants. The face with rounded eyes was more preferred than the face with spiky eyes regardless of the age of subjects. This tendency was different from the one observed in the shape condition in Experiment 1. Furthermore, the total time spent looking at the face tended to decrease as the subject’s age increased.

5.4 Discussion

Infants have preferences for visual stimuli that they are interested in and tend to look at these stimuli for a long time. It has been reported that this behavior changes from familiarity to novelty (Fantz, 1963; Houston-Price and Nakai, 2004). This behavioral method was adopted for the current study. This technique of using infants' preferential looking behavior has often been used in studies using point-light displays (PLDs) that depict biological motion (Christie and Slaughter, 2010; Slaughter et al., 2002; Zieber et al., 2005). In one study, infants looked at familiar items for a longer time (Sanefuji et al., 2008). In this study, the subjects were infants who could walk without support and infants who could crawl, and the PLDs of crawling and walking actions demonstrated by an adult were presented as stimuli. Infants who could crawl looked at the PLDs depicting crawling for a longer time than those depicting walking, and infants who could walk without support looked at the PLDs of walking for a longer time than those depicting crawling. This result suggests that infants prefer movement patterns that resemble their own motor repertoire.

Meanwhile, Morita et al. (2012) suggest that gaze tendency changes depending on the understanding of biomechanical constraints on arm movement. In this study, the length of time 9- and 12-month-old infants spent looking at animations of possible and impossible movements was measured. Although both 9- and 12-month-olds spent longer looking at the animation of impossible movement than the animation of possible movement, the 12-month-olds spent a significantly longer time looking at the animation of an impossible movement than did the 9-month-olds. These results suggest that the infants' visual attention was attracted to impossible movements due to their understanding of how arms can move.

Gaze tendency could also be influenced by individual differences. In one study, 8-month-old infants whose average motor skills were higher than the median looked at impossible movements longer than possible movements (Reid et al., 2005). This result suggests that gaze tendency was influenced by changes in understanding the movement of arms due to differences in individual motor skills. These differences were not limited to movement pattern processing. Another study that examined the association between frequency modulation and object size in infants found no difference in the gaze tendency in 4-month-old infants, but in 6-month-olds, the time spent on congruent conditions was longer. (Fernández-Prieto et al., 2015). In addition, in the study which investigated the connection between single syllables and object size in 4-month-old infants, the time spent attending to matching conditions was longer (Peña et al., 2011). In the study that examined the association between words and the roundness/sharpness of shapes in 4-month-old infants, the time spent on the incongruent condition was longer (Ozturk et al., 2013). These studies suggest that gaze tendency could change depending on the age of infants and the difficulty of the experiment.

In both the shape condition and the size condition in Experiment 1, when the sound stimulus was a tone burst, the looking time in the congruent condition was longer than that in the incongruent condition, regardless of the subject's age. When the sound stimulus was a Japanese phoneme, the gaze tendency differed between 9-month-old infants and 12- and 15-month-olds. In 12- and 15-month-old infants, the age at which many infants produce their first word, the looking time in the incongruent condition may be attributed to the change in the relationship between sound symbolism and speech production time. Since sound symbolism changed around the time of speech production, it was considered that the looking time in the incongruent condition was longer in 12- and 15-month-old

infants, when many infants produce their first word. Similarly, sound symbolism for tone bursts was suggested to have been acquired and changed around or before 9 months of age.

Previous studies using words with infants have used CVCV (consonant-vowel-consonant-vowel) words and CVCV-CVCV (repetition of CVCV) words (Ozturk et al., 2013; Asano et al., 2015; Imai et al., 2015). In Imai et al.'s (2015) study of 14-month-old infants, the connection between a two-mora word and a figure was examined, and it was found that the gaze time for the congruent condition was longer. In Experiment 2, no particular tendency was observed in the difference in looking time between the congruent condition and the incongruent condition, regardless of the subject's age. Since four-mora words were used as sound stimuli in Experiment 2, unlike the sound stimuli used in these previous studies and Experiment 1, it was considered that these sound stimuli are a more complex sound for infants. In addition, there may be a possibility that the impression recalled from the word was different between adults and infants, and that the impression of the shape expected by using the face as the visual stimulus could not be recalled. In spite of these problems, the results of Experiment 2 suggest that sound symbolism for words which have a higher number of morae has not yet been acquired at 15-month-olds. These issues should be verified through further experiments, such as reducing the number of morae in words and using the same visual stimuli as in Experiment 1.

Regarding the preference of visual stimuli, infants tended to look at the spiky shape longer in the shape condition of Experiment 1, but in Experiment 2, infants tended to look longer at the face with rounded eyes. Unlike Experiment 1, Experiment 2 displayed the different shapes as eye shapes. It has been reported that a few days after birth, infants look longer at the schematic of a face than at other shapes (Fantz, 1963). When the visual

stimulus was presented as an eye shape, the face with rounded eyes was less unnatural than the face with spiky eyes, and the characteristics of preferring a face-like shape were observed. In this study, sound symbolism during the language development was examined. These results suggest that there are several phases in the acquisition of sound symbolism. In other words, it was suggested that sound symbolism for phonemes is acquired at approximately 12 months of age, which is the speech production period, after sound symbolism for tone bursts is acquired approximately 9 months of age. Furthermore, it was suggested that sound symbolism for complex words is not yet acquired in 15-month-olds.

5.5 Summary

In this study, I measured infants' preferential looking behavior to stimuli, and two experiments were conducted to investigate multiple types of sound symbolism. In Experiment 1, tone bursts and Japanese phonemes were used as sound stimuli, and different shapes or sizes of figures were used as visual stimuli. In Experiment 2, nonsense words ("kinkiri," "chingai," "moudan," and "yumon") were used as sound stimuli, and two faces with rounded eyes or spiky eyes were used as visual stimuli. Subjects were 9-, 12-, and 15-month-old infants during the language development period. I measured the time spent looking at each visual stimulus while sound stimuli were presented. In Experiment 1, when the sound stimuli were tone bursts, the looking time in the incongruent condition was longer than that in the congruent condition, regardless of infant age. When the sound stimuli were Japanese phonemes, the looking time was differed according to infant age. In 9-month-olds, the looking time in the congruent condition was longer than that in the incongruent condition. However, among 12- and 15-month-olds, the looking time in the incongruent condition was longer than that in the congruent condition. In Experiment 2, the tendency for sound symbolism was not observed. These results suggest that sound symbolism for phonemes is acquired at approximately 12 months of age, which is the speech production period, after sound symbolism for tone bursts is acquired at approximately 9 months of age. Furthermore, it was suggested that sound symbolism for complex words is not yet acquired in 15-month-olds.

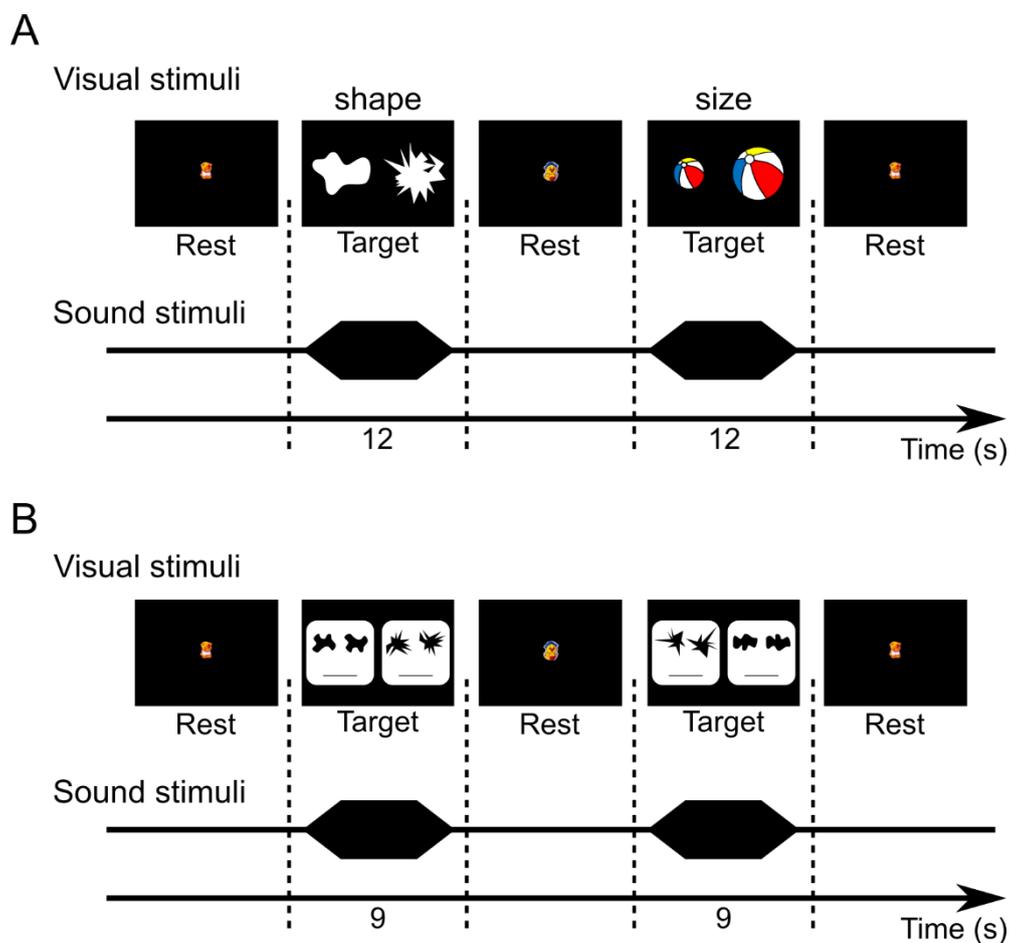


Figure 5.1 Illustration of the experimental design in Experiment 1 and Experiment 2. **(A)** Illustration of the experimental design in Experiment 1. The duration of one video was 12 s. All videos were presented randomly. Experiment 1 was divided into two sessions (eight videos in each session). **(B)** Illustration of the experimental design in Experiment 2. The duration of one video was 9 s. All videos were presented randomly. In order to acquire the attention of infants, videos that did not affect the experiment were presented before and after the experiment and between each stimulus, in Experiment 1 and 2 both.

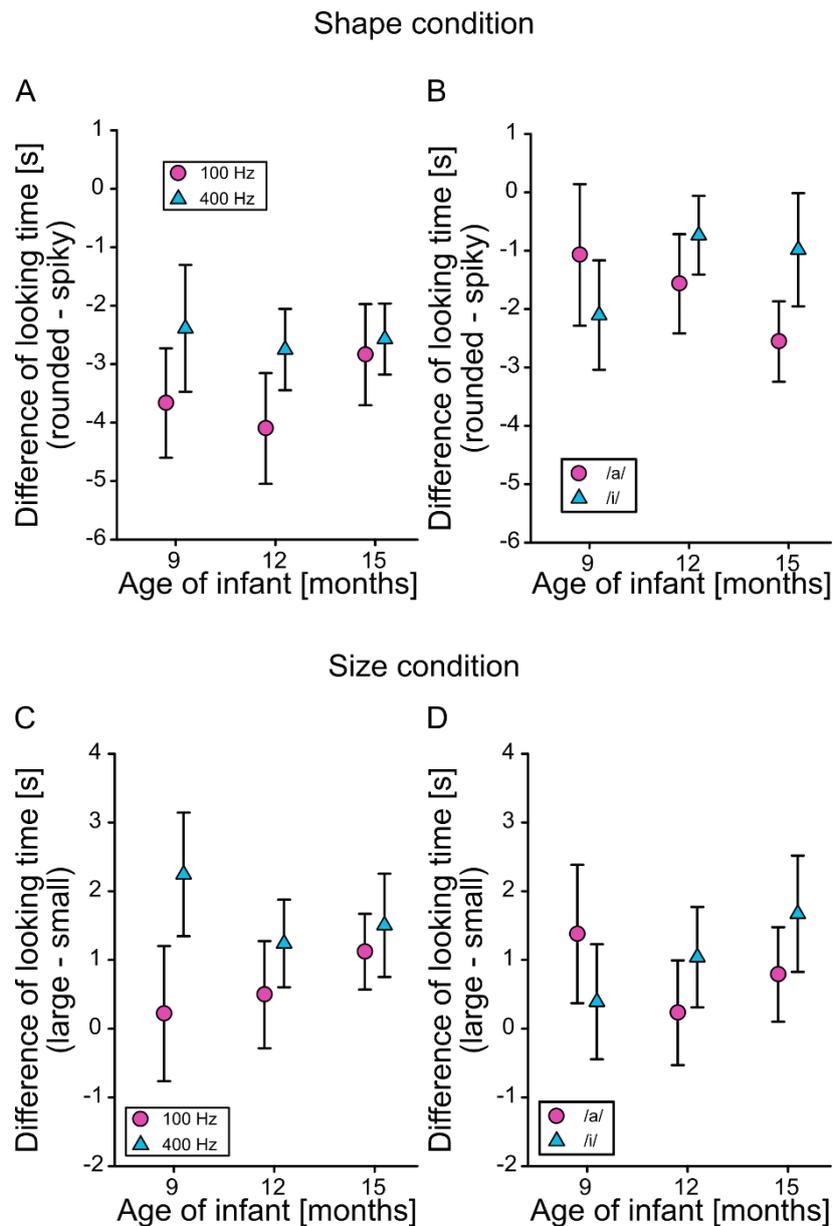


Figure 5.2 The differences in looking time in each condition.

(A) The sound stimuli were tone bursts in the shape condition.

(B) The sound stimuli were Japanese phonemes in the shape condition.

(C) The sound stimuli were tone bursts in the size condition.

(D) The sound stimuli were Japanese phonemes in the size condition.

The vertical axis represents the difference in looking time and the horizontal axis represents age of infants. Error bars represent the standard errors of the means. In **(A)** and **(C)**, the looking time in the incongruent condition was longer than that in the congruent condition, regardless of infant age. In **(B)** and **(D)**, the looking time in the congruent condition was longer than that in the incongruent condition in 9-month-olds. However, in 12- and 15-month-olds, the looking time in the incongruent condition was longer than that in the congruent condition.

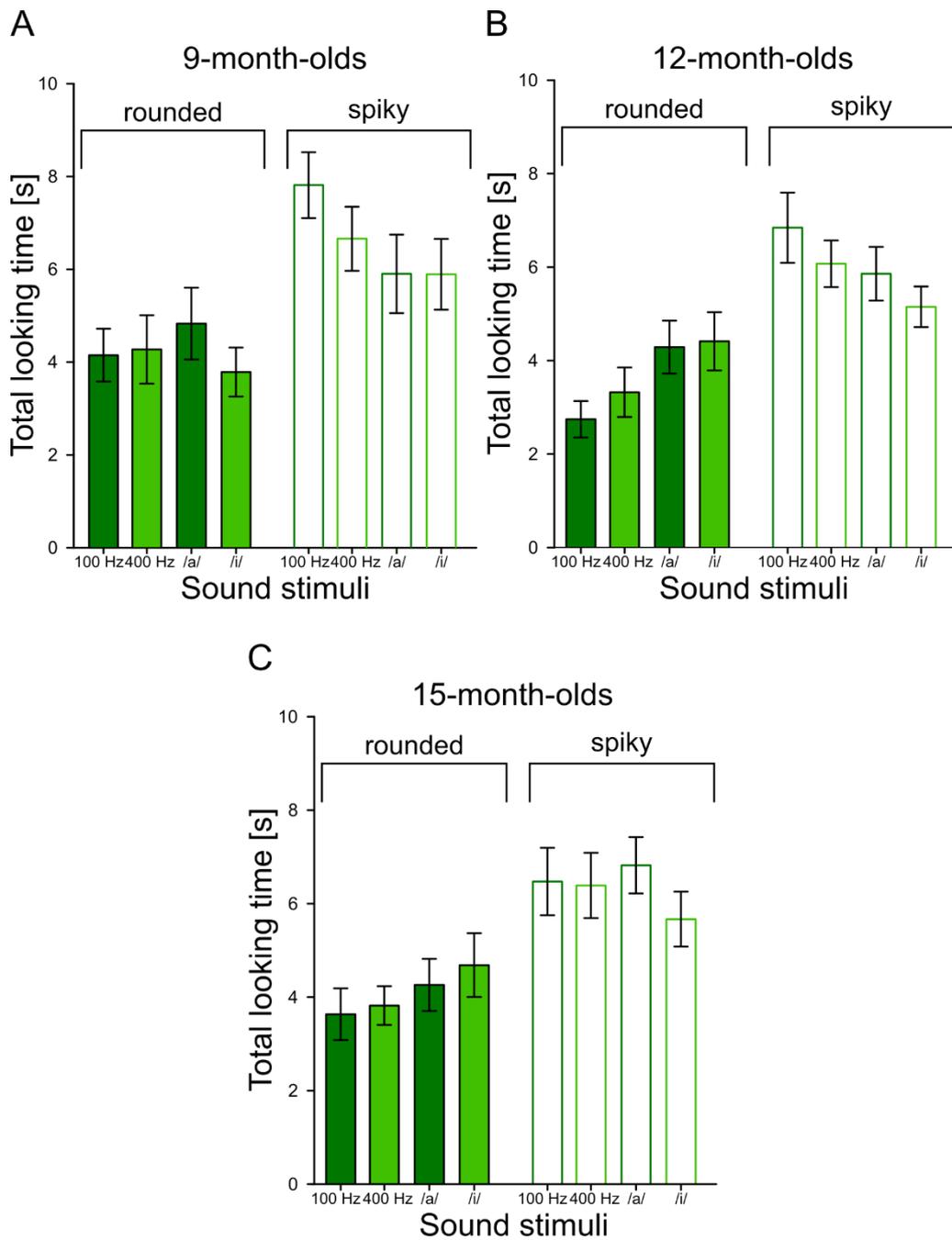


Figure 5.3 The looking time in each condition in the shape condition.

(A) The looking time in 9-month-olds.

(B) The looking time in 12-month-olds.

(C) The looking time in 15-month-olds.

The vertical axis represents the total looking time and the horizontal axis represents sound stimuli. Error bars represent the standard errors of the means. The looking time was longer for spiky shapes than for rounded shapes, regardless of infant age.

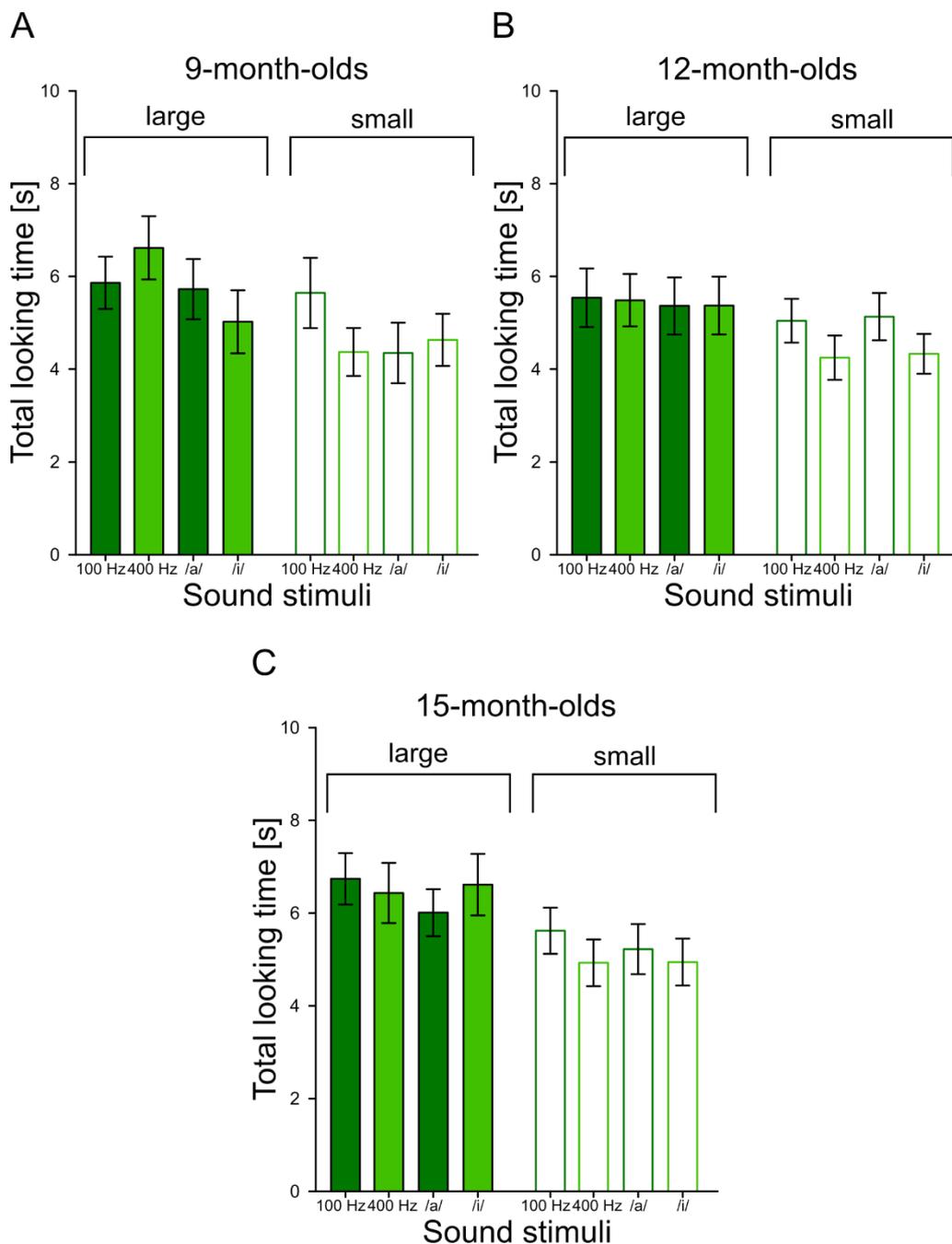


Figure 5.4 The looking time in each condition in the size condition.

(A) The looking time in 9-month-olds.

(B) The looking time in 12-month-olds.

(C) The looking time in 15-month-olds.

The vertical axis represents the total looking time and the horizontal axis represents sound stimuli. Error bars represent the standard errors of the means. When the sound stimulus was the same sound, the looking time was longer for the large ball compared to the small ball, regardless of infant age.

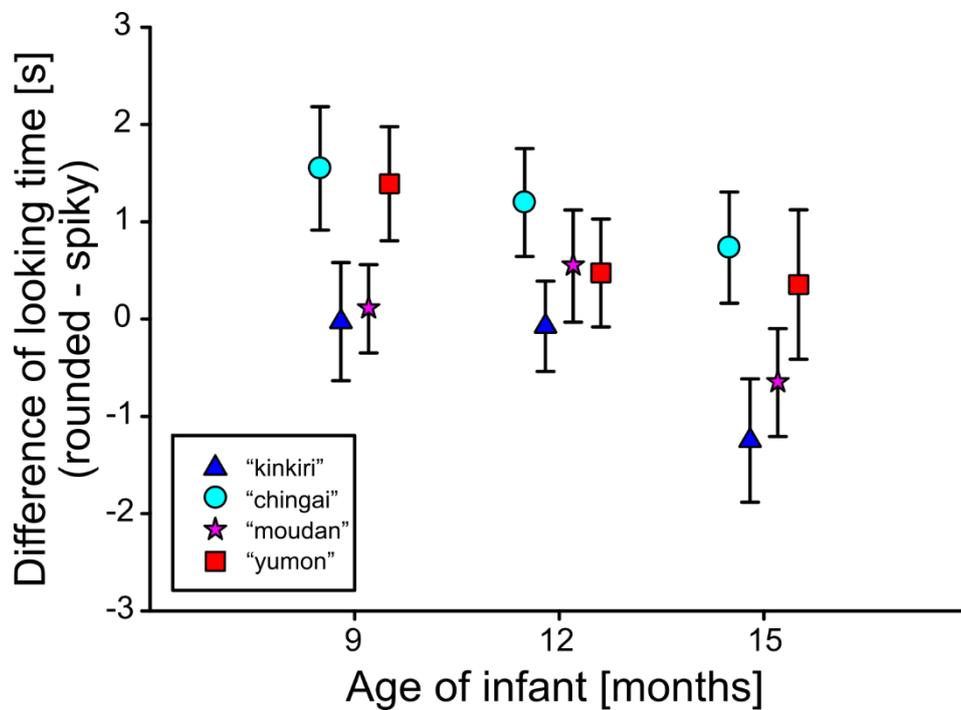


Figure 5.5 The differences in looking time in each condition. This figure shows the differences in looking time between the face with rounded eyes and the face with spiky eyes in Experiment 2. The vertical axis represents the difference in looking time and the horizontal axis represents age of infants. Error bars represent the standard errors of the means. When the word was “chingai,” the looking time at the face with rounded eyes was the longest, regardless of the age of subjects. When the word was “kinkiri,” there was a tendency for the looking time at the face with spiky eyes to be the longest, regardless of the age of subjects.

CHAPTER 6.

Change of sound symbolism during the language development: Infants' preferential reaching behavior

6.1 Introduction

Many object characteristics, such as complexity, size, and familiarity, can influence infants' looking preferences (Slater et al., 1990; Christie and Slaughter, 2010; Slaughter et al., 2002; Zieber et al., 2005). For example, newborns prefer to look at shapes that resemble a face (Fantz, 1963). As infants grow older and their visual abilities develop, other factors such as pattern complexity gain more importance in determining visual preference. While 3-week-old infants preferred to look at checkerboards made up of four large squares, by 8 weeks infants started to prefer patterns of intermediate complexity (made up of 64 squares), and by 14 weeks, infants preferred to look at patterns with the highest complexity (made up of 576 squares; Brennan et al., 1966). However, looking preference only suggests that infants can discriminate between objects.

Infants' behavior of reaching for objects that are interesting or familiar to them has been previously observed (Shinsky and Munakata, 2010; Newman et al., 2001; Fagard, 2000; Gurteen et al., 2011; Libertus et al., 2013). Shinsky and Munakata (2010) investigated the transition of preference from familiarity to novelty in infants. When objects were visible, 7- and 11-month-old infants showed a robust preference for novelty. However, when objects were hidden in the dark, 11-month-olds maintained

their novelty preference, in contrast to 7-month-olds who reversed to a familiarity preference. This result suggests that preferences change with development. One study examined the effect of object size on reaching preferences in infants aged 5 to 15 months (Newman et al., 2001). In this study, pairs of cylindrical objects 1 cm, 2.5 cm, or 6 cm in diameter were presented to infants, and the visual and reaching preferences were recorded. Between the ages of 5 months and 8.5 months, infants did not demonstrate a clear reaching preference for either smaller or larger objects. In contrast, infants aged 8.5 to 12 months were significantly more likely to reach for smaller, more graspable objects. These results show that perceived reachability influences the actual reaching behavior of infants for objects. In the current study, I measured the active selection behavior of infants using this characteristic.

In my previous study, I measured infants' looking time at two differently-shaped visual stimuli (i.e., rounded and spiky shapes). In 9-month-old infants, when the sound stimuli were Japanese phonemes, the looking time in the congruent condition (where the impression conveyed by the visual stimuli and sound stimuli matched) was longer than that in the incongruent condition (where the impressions did not match). On the contrary, in 12- and 15-month-old infants, the looking time in the incongruent condition was longer than that in the congruent condition. These results suggest that sound symbolism changed between 12 and 15 months of age, i.e., the time of speech production.

In this study, I investigated whether the preferential reaching behavior of infants was similar to infants' preferential looking time. If sound symbolism changed between 12 and 15 months of age, then infants' preferential reaching behavior could also change.

6.2 Materials and Methods

6.2.1 Subjects

Twelve 9-month-old infants (range, 8 months 24 days to 9 months 7 days; six boys, six girls), eleven 12-month-olds (range, 11 months 24 days to 12 months 7 days; four boys, seven girls), and ten 15-month-olds (range, 14 months 24 days to 15 months 7 days; four boys, six girls) participated in the study. All parents of infants who were participants gave written informed consent prior to participation, and they were native Japanese speakers.

6.2.2 Experimental apparatus

In this study, infants' behavior was observed while sitting on their parents' lap facing the testing table. Infants' behavior was recorded using two video cameras (SONY, FDR-AX60). One video camera was placed behind the experimenter to face the infant. Another camera recorded infants' behavior from the side. Objects were presented in pairs approximately 60 cm apart from each other on a rectangular tray. The sound stimulus was presented from a speaker (BOSE, computer music monitor) installed behind the experimenter.

6.2.3 Object stimuli

The objects were created with a 3D printer based on shapes used in my previous study. Two types of shape were used (two rounded and two spiky shapes; Fig. 6.1). These objects' sizes were 9 cm × 8.5 cm (range, 9 to 9.5 cm; 8 to 9.5 cm). These objects were shaken in tandem with the sound stimulus.

6.2.4 Sound stimuli

The sound stimuli were the Japanese phonemes /a/ and /i/. These phonemes were spoken by two native Japanese females. The duration of the sound was 250 ms, and the stimulus amplitude was adjusted to 74 (range 73 to 75) dB sound pressure level (SPL).

6.2.5 Procedure

The experiment began with a 15-s rest period. Then, two objects (rounded and spiky shapes) were presented clearly out of reach of the infant. At this time, these objects were shaken in tandem with the sound stimuli, which were presented four times. Subsequently, these objects were moved by the experimenter to a much closer position, clearly within the reach of the infant. The distance between objects was approximately 60 cm. At this time, the sound stimuli were again presented four times. One minute after these objects were moved closer to the infant, they took a rest (range, 12 s to 18 s) and started the next trial. Sounds unrelated to the experiment were presented during the rest period. When the infant dropped objects, their parent or experimenter picked them up and returned them to the testing table. There were a total of four trials to counterbalance the object position.

6.2.6 Analysis

In this study, I recorded the object that the infant touched first. The data from participants who failed to complete the four trials or did not touch the object were excluded. Five 9-month-old infants (two boys, three girls), seven 12-month-olds (two boys, five girls), and seven 15-month-olds (two boys, five girls) were excluded from the following analysis.

6.3 Results

I measured infants' reaching behavior by recording the object that each subject touched first (Table 6.1). The data from seven 9-month-old infants (four boys, three girls), four 12-month-olds (two boys, four girls), and three 15-month-olds (two boys and one girl) were analyzed. The data from two 9-month-old infants (subject 8 and subject 14), one 12-month-old (subject 7), and two 15-month-olds (subject 2 and subject 3) indicated a solitary preference for a particular position (left or right) in all trials. On the contrary, there were some participants (three 9-month-old infants: subject 9, subject 11, and subject 13; two 12-month-olds: subject 5 and subject 6) whose preference (left or right) was consistent in three of four trials. Therefore, I analyzed four videos (two 9-month-old infants: subject 10 and subject 12; one 12-month-old: subject 4; and one 15-month-old: subject 1).

Evidence from my previous study and other studies indicate that /a/ conveys an impression of roundness and /i/ conveys an impression of spikiness (Knoeferle et al. 2017, Fort et al. 2015). I defined the congruent and incongruent conditions as follows. In the congruent condition, the object was consistent with the reaction to the sound (i.e., the rounded object was presented with /a/). In the incongruent condition, the object was inconsistent with the reaction to the sound (i.e., the rounded object was presented with /i/).

Subjects 4, 10, and 12 selected the congruent object. Meanwhile, the 15-month-old infant (subject 1) selected the incongruent object.

6.4 Discussion

In my previous study, I measured the time spent looking at visual stimuli. In this study, two figures with different shapes (rounded and spiky shapes) were used as visual stimuli. The sound stimuli were presented in tandem with the visual stimuli being shaken. When the sound stimuli were Japanese phonemes, the looking time changed according to infant age. In 9-month-old infants, the looking time in the congruent conditions was longer than that in the incongruent conditions. In contrast, in 12- and 15-month-old infants, the looking time in the incongruent conditions was longer than that in the congruent condition. These results suggest that sound symbolism changes between 12 and 15 months of age, i.e., the time of speech production.

This experiment observed an active choice, unlike my previous study. Subjects 4, 10, and 12 selected the congruent object, while the 15-month-old infant (subject 1) selected the incongruent object. Not only the looking time but also the reaching behavior changed at approximately 12 months of age. This result suggests that sound symbolism may change at the time of speech production. Therefore, this result supports the findings of my previous study. However, further experiments should be conducted since there are little data.

6.5 Summary

In this study, I measured infants' reaching behavior towards objects. Objects were created with a 3D printer and based on two different types of shape (rounded and spiky shapes) used in my previous study. The sound stimuli were the Japanese phonemes /a/ and /i/, which were spoken by two native Japanese females. Objects were shaken left and right in tandem with the sound stimuli. Subjects were 9-, 12-, and 15-month-old infants during the language development period. I observed infants' preferential reaching behavior by recording the object that infants touched first. In this analysis, the data from subjects who failed to complete all trials or did not touch the object were excluded. Three 9- and 12-month-old infants selected the congruent objects across all trials. Only one 15-month-old infant selected the incongruent object across all trials. These results suggest that the preferential reaching behavior of infants changes at approximately 12 months of age, which is the time of speech production.

			Left; Rounded – Right; Spiky		Left; Spiky – Right; Rounded	
Subjects	Age [months]	Gender	/a/	/i/	/a/	/i/
Sub. 1	15	Male	Right	Left	Left	Right
Sub. 2	15	Male	Right	Right	Right	Right
Sub. 3	15	Female	Left	Left	Left	Left
Sub. 4	12	Male	Left	Right	Right	Left
Sub. 5	12	Male	Right	Left	Left	Left
Sub. 6	12	Female	Right	Left	Right	Right
Sub. 7	12	Female	Left	Left	Left	Left
Sub. 8	9	Male	Left	Left	Left	Left
Sub. 9	9	Male	Left	Right	Left	Left
Sub. 10	9	Male	Left	Right	Right	Left
Sub. 11	9	Male	Left	Left	Left	Right
Sub. 12	9	Female	Left	Right	Right	Left
Sub. 13	9	Female	Right	Right	Right	Left
Sub. 14	9	Female	Left	Left	Left	Left

Table 6.1 The first touch in infants' reaching behavior to objects in each subject.



Figure 6.1 Example of two differently-shaped toys used as stimuli.

CHAPTER 7.

Conclusions

In this dissertation, brain activity related to sound symbolism was measured to elucidate their cognitive and neural mechanism. In addition, sound symbolism during the language development was studied to investigate whether sound symbolism influences the language acquisition. In this chapter, I summarize the main results of the dissertation and discuss directions for future work.

7.1 Summaries of Main Results

7.1.1 Brain activity related to sound symbolism: Cross-modal effect of an aurally presented phoneme on judgment of size (Chapter 2)

In Chapter 2, the neural basis of sound symbolism was described. In this chapter, I described the effect of a phoneme (the basic unit of sound within a word) on the judgment of the size of a visual stimulus and the related brain activity. Subjects were required to perform two types of task, a comparison task and a control task. In the comparison task, subjects compared the visual sizes of standard and target stimuli. The target stimulus was either smaller or larger than the standard by $\pm 5\%$, $\pm 10\%$, or $\pm 20\%$ of its diameter, and was displayed to the subjects while they were listening to the sounds “bobo” or “pipi.” In the control task, subjects were asked to push the middle button

regardless of the visual size of the target stimuli. This task was designed to prevent subjects from ignoring the sound stimuli. According to the previous study on sound symbolism, it was identified that “bobo” is associated with a larger object and “pipi” is associated with a smaller object. Behavioral data showed that RT in the incongruent condition was longer than that in the congruent condition. This result suggests that in incongruent conditions, the task of determining visual size was the competitive condition for subjects. Furthermore, the RT difference between the incongruent condition and the congruent condition increased as the visual size difference between the target and the standard increased. The fMRI data showed that the bilateral ACCs were more activated in the incongruent condition than in the congruent condition, regardless of the differences in visual size. In addition, the left MTG and right STG were also activated in the incongruent condition. It was reported that these regions were related to semantic and phonetic processes. The result of this study suggests that the right STG is part of a brain network involved in processing conflict in phonemic sound symbolism, in addition to emotional prosodic information. This implies that symbolic matching between the target size and a phoneme is also processed in the MTG. These results demonstrate that the phonetic process influences sound symbolism.

7.1.2 Sound symbolism on synthetic speech continuum in judgment of size of visual stimulus (Chapter 3)

In Chapter 3, the effect of phoneme perception on sound symbolism was described. I created a phoneme continuum to serve as a stimulus and investigated whether perceptual difficulty changes sound symbolism. This phoneme continuum was a 10-step audio-morphed continua between pairs of naturally recorded speech sounds using “STRAIGHT” software. In Experiment 1, a stimulus continuum between /pi/ and /ba/ was generated, while a stimulus continuum between /i/ and /a/ was used in Experiment 2. Therefore, a total of 11 stimulus types (two originals and nine intermediates) were used in each experiment. The composite sound in which approximately 50 % of each sound component was mixed was perceptually very ambiguous and, for instance, was randomly interpreted as /pi/ or /ba/. The results of Experiments 1 and 2 show that when it was difficult to identify the sound stimulus as a phoneme, sound symbolism was not observed. On the other hand, when the sound stimulus was clearly identified as a phoneme, sound symbolism was observed. In other words, the RT in the incongruent condition was longer than in the congruent condition. This tendency was observed in both Experiment 1 and 2. This result suggests that the simple acoustic characteristic alone, which was manipulated as the morphing rate, was not the main contributor to sound symbolism. As such, phoneme perception could play a measurable role in sound symbolism.

7.1.3 Brain regions involved in sound symbolism and their relationship with phoneme perception (Chapter 4)

In Chapter 4, the relationship between speech perception and sound symbolism was described. In Experiment 1, I examined brain activity related to sound symbolism, and in Experiment 2, I examined the brain region related to speech perception. In Experiment 1, subjects were required to judge the difference in visual size between a standard and a target stimulus while listening to sound stimuli. Sound stimuli were single syllables (/bo/, /bi/, /po/ and /pi/), a noise and a click sound. In Experiment 2, subjects were asked to discriminate between the heard sounds. Only the Japanese phonemes that were used in Experiment 1 were used as sound stimuli in Experiment 2. Behavioral data showed that the RT underrun the incongruent condition was longer than in the congruent condition. That is, even when the sound stimulus was a single syllable, sound symbolism was still observed. Furthermore, when comparing the difference in RT between the incongruent and congruent condition, the influence of vowels was larger than that of consonants. The fMRI data for Experiment 1 showed that the left postcentral gyrus was more activated in the incongruent condition than in the congruent condition. This region was also activated during speech perception. In addition, other motor areas were also activated during speech perception. This result suggests that the incongruent condition was more difficult for perception and speech discrimination. In other words, sound symbolism may have boosted speech recognition and visual size determination. The values of the contrast estimate in the lower order area of the visual cortex were larger in the congruent condition than in the incongruent

condition. From this result it can be inferred that the first step of speech symbolism may be processed in the lower order areas of the visual cortex, such as V1 to V3. These results support that acoustic character is not the only factor involved in sound symbolism. Furthermore, it was suggested that the motor area related to speech perception influences sound symbolism.

7.1.4 Change of sound symbolism during the language development: Infants' preferential looking behavior (Chapter 5)

In Chapter 5, sound symbolism in infants was explored using infants' visual preferences. Two experiments were conducted to investigate multiple types of sound symbolism. In Experiment 1, tone bursts and Japanese phonemes were used as sound stimuli, and objects of different shapes or sizes were used as visual stimuli. In Experiment 2, non-sense words ("kinkiri," "chingai," "moudan," and "yumon") were used as sound stimuli, and two faces with rounded eyes or spiky eyes were used as visual stimuli. Subjects were 9-, 12-, and 15-month-old infants during the language development period. I measured the time spent looking at each visual stimulus while sound stimuli were presented. In Experiment 1, when the sound stimuli were tone bursts, the looking time in the incongruent condition was longer than in the congruent condition, regardless of infant age. On the other hand, when the sound stimuli were Japanese phonemes, the looking time changed depending on infant age. In 9-month-olds, the looking time in the congruent condition was longer than in the incongruent condition. However, 12- and 15-month-olds spent a longer

time looking in the incongruent condition than in the congruent condition. These tendencies were observed in both visual stimuli conditions (i.e., shape and size). These results suggest that sound symbolism for tone bursts is attained before the age of 9 months, while sound symbolism for Japanese phonemes changes at approximately 12 months old. In Experiment 2, sound symbolism was not observed. This result suggests that sound symbolism for words is not exhibited at 15 months of age.

7.1.5 Change of sound symbolism during the language development: Infants' preferential reaching behavior (Chapter 6)

In Chapter 6 described infants' preferential reaching behavior in regard to sound symbolism. In this study, two video cameras were used to record infants' reaching behavior towards objects. Four objects were created with a 3D printer based on two different types of shape (rounded and spiky shapes), as used in my previous study. The sound stimuli were the Japanese phonemes /a/ and /i/ which were spoken by two native Japanese females. The objects were shaken left and right in tandem with sound stimuli. This study investigated whether infants' reaching behavior was influenced by sound stimuli, which were presented in tandem with the object's movement. Participants were 9-, 12-, and 15-month-old infants during the language development period. I compared the objects that infants touched first. In this analysis, I excluded data from participants who failed to complete all trials or did not touch the object. The data from 10 participants showed a preference for a particular position in

all or most trials. Three 9- and 12-month-old infants selected the congruent objects across all trials. On the other hand, one 15-month-old infant selected the incongruent objects across all trials. These results suggest that the preferential reaching behavior of infants change at approximately 12 months of age, which is the time of speech production. This suggests that sound symbolism for Japanese phonemes changes at approximately 12 months old.

7.2 Future works

7.2.1 The effect of sound symbolism on language learning

In my study, sound symbolism was observed even in infants, suggesting that sound symbolism changes around the time of speech production. In other words, sound symbolism can affect language development. However, this result does not reveal whether sound symbolism contributes to language learning. In one study, unknown natural language ideophones were used to examine whether naive speakers could choose the correct translation of the word (Lockwood et al. 2016). The results showed that the subject was able to guess the correct translation of the unknown ideophone. This result suggests that sound symbolism contributed to identifying the meaning of the ideogram, but it was insufficient to clarify whether sound symbolism influences language learning. That is, in order to examine the relationship between language learning and sound symbolism, it is necessary to compare the learning of symbolic words and the learning of non-symbolic words. Therefore, in future research, the group that learns unknown words with symbolic sound

combinations should be compared with another group that learns unknown words with non-symbolic sound combinations. Furthermore, these learning effects should be considered at various ages.

7.2.2 Changes in sound symbolism

This study suggests that sound symbolism changes at the time of language development in infants, and that the change depends on the type of sound symbolism. However, the results of this study did not reveal when sound symbolism is acquired. When the sound stimuli were tone bursts, the preferential gaze tendency of infants to incongruent conditions was observed at 9 months old. If sound symbolism was not acquired by this point, there would be no difference in gaze preference between congruent and incongruent conditions. Therefore, sound symbolism in infants 9 months old and younger should be measured. On the other hand, when the sound stimuli were words, sound symbolism was not observed, regardless of infant age. The experimental paradigm used in this study may not have been optimal to investigate sound symbolism for words. Future research should improve this experimental paradigm and measure sound symbols at various stages of infancy.

7.3 Final remarks

This dissertation describes the brain activity that contributes to sound symbolism and sound symbolism in infants during the language development.

Consequently, the RT in the incongruent condition was longer than in the congruent condition, it was found that sound symbolism was observed even when sound stimuli were presented aurally. The ACC was more activated in the incongruent condition than congruent condition. The activation of the dorsal region of the ACC in the incongruent condition may reflect the Stroop-like interference between the phoneme and the visual magnitude of the stimulus. In addition, the left MTG and right STG were activated in the incongruent condition. These results suggest that the right STG is part of a brain network involved in processing conflict in phonemic sound symbolism in addition to emotional prosodic information. These results also imply that symbolically matching sounds between target size and phonemes is processed in the MTG. Thus, the possibility that phonological processing contributes to phonetic symbols is supported. This dissertation then described the relationship between phoneme perception and sound symbolism. Consequently, it was shown that the left postcentral gyrus, which is related to phoneme perception was more activated in the incongruent condition. This result suggests that tasks in the incongruent condition were more difficult for perception and speech discrimination. Therefore, this dissertation shows that simple acoustic characteristics are not the only contributors to sound symbolism. Indeed, speech perception and the motor areas related to speech perception also contribute to sound symbolism. Furthermore, this dissertation explained that sound symbolism changes in infants during the language development and could contribute to language acquisition itself. In all, the results suggest that motor cortex involved in phoneme articulation contributes

to sound symbolism and that effect of sound symbolism and language develop through mutually interacting with each other. The findings in this study provide new insights into research aimed at elucidating the mechanisms of language development and language acquisition.

REFERENCES

- Asano, M. *et al.* Sound symbolism scaffolds language development in preverbal infants. *Cortex* **63**, 196-205 (2015).
- Aso, T., Jiang, G., Urayama, S. & Fukuyama, H. A resilient, non-neuronal source of the spatiotemporal lag structure detected by BOLD signal-based blood flow tracking. *Front. Neurosci.* **11**, 256 (2017).
- Auracher, J., Albers, S., Zhai, Y., Gareeva, G. & Stavniychuk, T. P is for happiness, N is for sadness: Universals in sound iconicity to detect emotions in poetry. *Discourse Process.* **48**, 1-25 (2010).
- Bernstein, I. H. & Edelman, B. A. Effects of some variations in auditory input upon visual choice reaction time. *J. Exp. Psychol.* **87**, 241-247 (1971).
- Bond, B. & Stevens, S. S. Cross-modality matching of brightness to loudness by 5-year-olds. *Percept. Psychophys.* **6**, 337-339 (1969).
- Bosma, J. F. Anatomic and physiologic development of the speech apparatus. Tower, D. B. (Ed.), *The nervous system, vol. 3: Human communication and its disorders.* Raven Press, New York, 469-480 (1975).
- Brennan, W. M., Ames, E. & Moore, R. W. Age differences in infants' attention to pattern of different complexities. *Science* **151**, 354-356 (1966).
- Chiou, R. & Rich, A. N. Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception* **41**, 339-353 (2012).
- Christie, T. & Slaughter, V. Movement contributes to infants' recognition of the human form. *Cognition* **114**, 329-337 (2010).

- Correia, J. M., Jansma, B. M. B. & Bonte, M. Decoding articulatory features from fMRI responses in dorsal speech regions. *J. Neurosci.* **35**, 15015-15025 (2015).
- de Saussure, F. *Course in general linguistics* (Open Court, 1983).
- Ehrsson, H. H., Geyer, S. & Naito, E. Imagery of voluntary movement of fingers, toes, and tongue activates corresponding Body-Part-Specific motor representations. *J. Neurophysiol.* **90**, 3304-3316 (2003).
- Erdoğan, S. B., Tong, Y., Hocke, L. M., Lindsey, K. P. & Frederick, B. d. Correcting for blood arrival time in global mean regression enhances functional connectivity analysis of resting state fMRI-BOLD signals. *Front. Hum. Neurosci.* **10**, 311 (2016).
- Esterman, M., Tamber-Rosenau, B. J., Chiu, Y. C. & Yantis, S. Avoiding non-independence in fMRI data analysis: Leave one subject out. *Neuroimage* **50**, 572-576 (2010).
- Fadiga, L., Craighero, L., Buccino, G. & Rizzolatti, G. Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *Eur. J. Neurosci.* **15**, 399-402 (2002).
- Fagard, J. Linked proximal and distal changes in the reaching behavior of 5- to 12-month-old human infants grasping objects of different sizes. *Infant Behav. Dev.* **23**, 317-329 (2000).
- Fang, F., Boyaci, H., Kersten, D. & Murray, S. O. Attention-dependent representation of a size illusion in human V1. *Curr. Biol.* **18**, 1707-1712 (2008).
- Fantz, R. L. Pattern vision in newborn infants. *Science* **140**, 296-297 (1963).
- Fernandez-Prieto, I., Nabarra, J. & Pons, F. How big this sound? Crossmodal association between pitch and size in infants. *Infant Behav. Dev.* **38**, 77-81 (2015).
- Fort, M., Martin, A. & Peperkamp, S. Consonants are more important than vowels in the

- Bouba-kiki effect. *Lang. Speech.* **58**, 247-266 (2015).
- Freed, D. J. Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *J. Acoust. Soc. Am.* **87**, 311-322 (1990).
- Gallace, A. & Spence, C. Multisensory synesthetic interactions in the speeded classification of visual size. *Percept. Psychophys.* **68**, 1191-203 (2006).
- Grabski, K. *et al.* Functional MRI assessment of orofacial articulators: Neural correlates of lip, jaw, larynx, and tongue movements. *Hum. Brain Mapp.* **33**, 2306-2321 (2012).
- Gurteen, P. M., Horne, P. J. & Mihela, E. Rapid word learning in 13- and 17-month-olds in a naturalistic two-word procedure: Looking versus reaching measures. *J. Exp. Child Psychol.* **109**, 201-217 (2011).
- Hickok, G. & Poeppel, D. Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* **4**, 131-138 (2000).
- Hirata, S., Ukita, J. & Kita, S. Compatibility between pronunciation of voiced / voiceless consonants and brightness of visual stimuli. *Cogn. Stud.* **18**, 470-476 (2011).
- Houston-Price, C. & Nakai, S. Distinguishing novelty and familiarity effects in infant preference procedures. *Infant Child Dev.* **13**, 341-348 (2004).
- Imai, M., Kita, S., Nagumo, M. & Okada, H. Sound symbolism facilitates early verb learning. *Cognition* **109**, 54-65 (2008).
- Imai, M. *et al.* Sound symbolism facilitates word learning in 14-month-old. *PLoS One* **10**, e0116494 (2015).
- Itagaki, S. & Kobayasi, I. K. Cross-modal effect of an aurally presented phoneme on judgment of a visual object's size. *Harris Sci. Rev. Doshisha Univ.* **57**, 239-243 (2017).

- Itagaki, S., Murai, S. & Kobayasi, I. K. Brain activity related to sound symbolism: Cross-modal effect of an aurally presented phoneme on judgment of size. *Sci. Rep.* **9**, 7017 (2019).
- Kanero, J., Imai, M., Okuda, J., Okada, H. & Matsuda, T. How sound symbolism is processed in the brain: A study on Japanese mimetic words. *PLoS One* **9**, e97905 (2014).
- Kantartzis, K., Imai, M. & Kita, S. Japanese sound-symbolism facilitates word learning in English-speaking children, *Cognitive Sci.* **35**, 575-586 (2011).
- Kawahara, H., Masuda-Katsuse, I. & de Cheveigné, A. Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Commun.* **27**, 187-207 (1999).
- Kent, R. D. Articulatory-acoustic perspectives on speech development. Stark, R. E. (Ed.), *Language Behavior in Infancy and Early Childhood*. Elsevier Press, New York 105-126 (1981).
- Kent, R. D. The biology of phonological development. Ferguson C. A., Menn. L., Stoel-Gammon, C. (Eds.), *Phonological development: Models, research, implications*. York Press, Timonium, MD 65-90 (1992).
- Knoeferle, K., Li, J., Maggioni, E. & Spence, C. What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Sci. Rep.* **7**, 5562 (2017).
- Köhler, W. *Gestalt psychology 2nd ed.* (Liveright, 1947).
- Konkle, T. & Oliva, A. A real-world size organization of object responses in occipitotemporal cortex. *Neuron* **74**, 1114-1124 (2012).

- Kourtzi, Z. & Connor, C. E. Neural representations for object perception: Structure, category, and adaptive coding. *Annu. Rev. Neurosci.* **34**, 45-67 (2011).
- Kovic, V., Plunkett, K. & Westermann, G. The shape of words in the brain. *Cognition* **114**, 19-28 (2010).
- Kuhl, P. K. & Meltzoff, A. N. Infant vocalizations in response to speech: Vocal imitation and developmental change. *J. Acoust. Soc. Am.* **100**, 2425-2438 (1996).
- Kuhl, P. K. Early language acquisition: Cracking the speech code. *Nat. Rev. Neurosci.* **5**, 831-843 (2004).
- Lieberman, A. M. & Mattingly, I. G. The motor theory of speech perception revised. *Cognition* **21**, 1-36 (1985).
- Lieberman, A. M. & Whalen, D. H. On the relation of speech to language. *Trends Cogn. Sci.* **4**, 187-196 (2000).
- Libertus, K. *et al.* Size matters: How age and reaching experiences shape infants' preferences for different sized objects. *Infant Behav. Dev.* **36**, 189-198 (2013).
- Lockwood, G. & Dingemanse, M. Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Front. Psychol.* **6**, 1246 (2015).
- Lockwood, G., Dingemanse, M. & Hagoort, P. Sound-symbolism boosts novel word learning. *J. Exp. Psychol. Learn.* **42**, 1274-1281 (2016).
- Ludwig, V. U., Adachi, I. & Matsuzawa, T. Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (*Pan troglodytes*) and humans. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 20661-20665 (2011).
- Marks, L. E. On associations of light and sound: The mediation of brightness, pitch, and loudness. *Am. J. Psychol.* **87**, 173-188 (1974).

- Marks, L. E. On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *J. Exp. Psychol. Human.* **13**, 384-394 (1987).
- Maurer, D., Pathman, T. & Mondloch, C. J. The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Sci.* **9**, 316-322 (2006).
- McGurk, H. & MacDonald, J. Hearing lips and seeing voices. *Nature* **264**, 746-748 (1976).
- Morita, T. *et al.* Infant and adult perceptions of possible and impossible body movements: An eye-tracking study. *J. Exp. Child Psychol.* **113**, 401-414 (2012).
- Murray, S. O., Boyaci, H. & Kersten, D. The representation of perceived angular size in human primary visual cortex. *Nat. Neurosci.*, **9**, 429-434 (2006).
- Nagatani, Y., Tachibana, R., Sakaguchi, T. & Hosoi, H. Loudness calibration of monosyllabic speech sounds in FW03. *J. Acoust. Soc. Jpn.* **64**, 647-649 (2008).
- Newman, S. S. Further experiments in phonetic symbolism. *Am. J. Psychol.* **45**, 53-75 (1933).
- Newman, C., Atkinson, J. & Braddick, O. The development of reaching and looking preferences in infants to objects of different size. *Dev. Psychol.* **37**, 561-572 (2001).
- Nielsen, A. & Rendall, D. Parsing the role of consonants versus vowels in the classic takete-maluma phenomenon. *Can. J. Exp. Psychol.* **67**, 153-163 (2013).
- Noppeney, U., Josephs, O., Hocking, J., Price, C. J. & Friston, K. J. The effect of prior visual information on recognition of speech and sounds. *Cereb. Cortex* **18**, 598-609 (2008).
- Nygaard, L. C., Cook, A. E. & Namy, L. L. Sound to meaning correspondences facilitate word learning. *Cognition* **112**, 181-186 (2009).
- Obleser, J. *et al.* Vowel sound extraction in anterior superior temporal cortex. *Hum. Brain Mapp.* **27**, 562-571 (2006).

- Ohtake, Y. & Haryu E. Investigation of the process underpinning vowel-size correspondence. *Jpn. Psychol. Res.* **55**, 390-399 (2013).
- Ogura, T. & Watamaki, T. MacArthur Communicative Development Inventories: User's Guide and Technical Manual. *Kyoto: Kyoto International Social Welfare Exchange Centre* (2004).
- Oller, D.K. & Lynch, M. P. Infant vocalizations and innovations in infraphonology: Toward a broader theory of development and disorders. Ferguson C. A., Menn. L., Stoel-Gammon, C. (Eds.), *Phonological development: Models, research, implications*. York Press, Timonium, MD 509-536 (1992).
- Osaka, N. *et al.* An emotion-based facial expression word activates laughter module in the human brain: A functional magnetic resonance imaging study. *Neurosci. Lett.* **340**, 127-130 (2003).
- Osaka, N., Osaka, M., Morishita, M., Kondo, H. & Fukuyama, H. A word expressing affective pain activates the anterior cingulate cortex in the human brain: An fMRI study. *Behav. Brain Res.* **153**, 123-127 (2004).
- Osaka, N. Walk-related mimic word activates the extrastriate visual cortex in the human brain: An fMRI study. *Behav. Brain Res.* **198**, 186-189 (2009).
- Ozturk, O., Krehm, M. & Vouloumanos, A. Sound symbolism in infancy: Evidence for sound-shape cross-modal correspondences in 4-month-olds. *J. Exp. Child Psychol.* **114**, 173-186 (2013).
- Pardo, J. V., Pardo, P. J., Janer, K. W. & Raichle, M. E. The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 256-259 (1990).
- Parise, C. & Spence, C. Synesthetic congruency modulates the temporal ventriloquism

- effet. *Neurosci. Lett.* **442**, 257-261 (2008).
- Parise, C. V. & Pavani, F. Evidence of sound symbolism in simple vocalizations. *Exp. Brain. Res.* **214**, 373-380 (2011).
- Parise, C. V. & Spence, C. Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Exp. Brain. Res.* **220**, 319-333 (2012).
- Peña, M., Mehler, J. & Nespors, M. The role of audiovisual processing in early conceptual development. *Psychol. Sci.* **22**, 1419-1421 (2011).
- Pooresmaeili, A., Arrighi, R., Biagi, L. & Morrone, M. C. Blood oxygen level-dependent activation of the primary visual cortex predicts size adaptation illusion. *J. Neurosci.* **33**, 15999-16008 (2013).
- Price, C. J. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* **62**, 816-847 (2012).
- Ramachandran, V. S. & Hubbard, E. M. Synaesthesia-A window into perception, thought and language. *J. Conscious. Stud.* **8**, 3-34 (2001).
- Reid, V. M., Belsky, J. & Johnson, M. H. Infant perception of human action: Toward a developmental cognitive neuroscience of individual differences. *Cognition, Brain, Behavior* **4**, 193-210 (2005).
- Revill, K. P., Namy, L. L., DeFife, L. C. & Nygaard, L. C. Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI. *Brain Lang.* **128**, 18-24 (2014).
- Roelofs, A., Turenout, M. Van & Coles, M. G. H. Anterior cingulate cortex activity can be independent of response conflict in Stroop-like tasks. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 13884-13889 (2006).

- Sanefuji, W., Ohgami, H. & Hashiya, K. Detection of the relevant type of locomotion in infancy: Crawlers versus walkers. *Infant Behav. Dev.* **31**, 624-628 (2008).
- Sapir, E. A study in phonetic symbolism. *J. Exp. Psychol.* **12**, 225-239 (1929).
- Schneider, T. R., Debener, S., Oostenveld, R. & Engel, A. K. Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *Neuroimage* **42**, 1244-1254 (2008).
- Schwarzkopf, D. S. & Rees, G. Subjective size perception depends on central visual cortical magnification in human V1. *PLoS One* **8**, e60550 (2013).
- Scott, S. K. & Johnsrude, I. S. The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* **26**, 100-107 (2003).
- Shinohara, K. & Kawahara, S. A cross-linguistic study of sound symbolism: The images of size. *Proceedings of BLS 36*. Berkeley: Berkeley Linguistic Society. 396-410 (2010).
- Shinohara, K., Yamauchi, N., Kawahara, S. & Tanaka, H. Takete and Maluma in action: A cross-modal relationship between gestures and sounds. *PLoS One* **11**, e0163525 (2016).
- Shinsky, J. L. & Munakata, Y. Something old, something new: A developmental transition from familiarity to novelty preferences with hidden objects. *Developmental Sci.* **13**, 378-384 (2010).
- Slater, A., Mattock, A. & Brown, E. Size constancy at birth – newborn – infants responses to retinal and real size. *J. Exp. Child Psychol.* **49**, 314-322 (1990).
- Slaughter, V., Herson, M. & Sim, S. Development of preferences for the human body shape in infancy. *Cognition* **85**, B71-B87 (2002).
- Smith, S. M. *et al.* Advances in functional and structural MR image analysis and

- implementation as FSL. *Neuroimage* **23**, S208-S219 (2004).
- Spence, C. Crossmodal correspondences: A tutorial review. *Atten. Percept. Psycho.* **73**, 971-995 (2011).
- Stoel-Gammon, C. Prelinguistic vocal development: Measurement and predictions. Ferguson C. A., Menn. L., Stoel-Gammon, C. (Eds.), *Phonological development: Models, research, implications*. York Press, Timonium, MD 439-456 (1992).
- Tong, Y. *et al.* Evaluating the effects of systemic low frequency oscillations measured in the periphery on the independent component analysis results of resting state networks. *Neuroimage* **76**, 202-215 (2013).
- Ungerleider, L. G. & Bell, A. H. Uncovering the visual “alphabet”: Advances in our understanding of object perception. *Vision Res.* **51**, 782-799 (2011).
- Walker, P. & Smith, S. Stroop interference based on the multimodal correlates of haptic size and auditory pitch. *Perception* **14**, 729-736 (1985).
- Walker, P. *et al.* Preverbal infants’ sensitivity to synaesthetic cross-modality correspondences. *Psychol. Sci.* **21**, 21-25 (2010).
- Watson, R. *et al.* Dissociating task difficulty from incongruence in face-voice emotion integration. *Front. Hum. Neurosci.* **7**, 744 (2013).
- Westbury, C. Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain Lang.* **93**, 10-19 (2005).
- Wittfoth, M. *et al.* On emotional conflict: Interference resolution of happy and angry prosody reveals valence-specific effects. *Cereb. Cortex* **20**, 383-392 (2010).
- Zieber, N., Kangas, A., Hock, A. & Bhatt, R. S. Body structure perception in infancy. *Infancy* **20**, 1-17 (2015).

CURRICULUM VITAE

Sachi Itagaki

Education:

M.Sc., Engineering, Doshisha University, Kyotanabe, Kyoto, Japan, 2017

B.Sc., Engineering, Doshisha University, Kyotanabe, Kyoto, Japan, 2015

Grants and Fellowship:

Research Fellowships of Japan Society for the Promotion of Science for
Young scientists (DC1; 04/2017-03/2020)

November 2019