



Neural mechanisms of spatial processing and learning in human-to-human communication: fMRI studies.

カラン, 明子

(Degree)

博士 (学術)

(Date of Degree)

2015-07-17

(Date of Publication)

2016-07-01

(Resource Type)

doctoral thesis

(Report Number)

乙第3288号

(URL)

<https://hdl.handle.net/20.500.14094/D2003288>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



博士論文

Neural mechanisms of spatial processing and
learning in human-to-human communication:
fMRI studies.

(ヒューマンコミュニケーションにおける空間処理と学習
の脳内メカニズム：脳機能画像研究)

平成27年7月

カラン明子

Acknowledgements

I would like to express my gratitude to all those who have supported me to accomplish the research presented in this dissertation. First of all, I would like to thank Dr. Ryoko Hayashi for giving me the opportunity to pursue a Ph. D., for organizing the dissertation committee, and for providing proper guidance. I would like to give thanks to Dr. Eriko Matsumoto for critical reading of my dissertation and for giving me constructive comments. I am very fortunate to be able to work with Dr. Reiko Akahane-Yamada who is a leading researcher in foreign language learning and would like to express my sincere gratitude for her valuable opinions and encouragement. I would also like to show my deep gratitude to Dr. Shinobu Masaki who educated me to be a researcher and introduced me to fMRI research. I am very grateful to Dr. Hiroshi Ando for allowing me to do fMRI experiments and encouraging me to pursue a Ph. D. Finally, I would like to express my sincere appreciation to Dr. Hiroaki Kato for inspiring and providing knowledgeable advice.

Contents

Acknowledgements

Chapter 1. Introduction	1
1.1 Why we study human-to-human communication with fMRI?	1
1.2 About fMRI studies	3
1.3 Purposes of the study	4
1.3.1 Spatial processing in communication: Integration of audiovisual spatial information in speech perception.	4
1.3.2 Learning in communication: Neural response differences between native and non-native linguistic rules.....	6
1.4 Outline of the dissertation	7

Spatial processing in communication: Integration of audiovisual spatial information	10
--	-----------

Chapter 2. Brain regions involved with auditory spatial processing in the real world	11
2.1 Introduction	11
2.2 Materials and Methods.....	15
2.2.1 Participants	15
2.2.2 Stimuli and procedure.....	16
2.2.3 MRI data acquisition and preprocessing	18
2.2.4 fMRI data analysis	19
2.3 Results	20
2.3.1 Behavioral data for experiment 1	20
2.3.2 fMRI data for experiment 1	20

2.3.3 Behavioral data for Experiment 2	21
2.3.4 fMRI data for experiment 2	21
2.4 Discussion	24
2.5 Conclusion	29
 Chapter 3. Neural basis of audiovisual spatial integration in speech perception	 30
3.1 Introduction	30
3.2 Materials and Methods.....	34
3.2.1 Participants	34
3.2.2 Stimuli	35
3.2.3 Procedure	37
3.2.4 MRI data acquisition and preprocessing	39
3.2.5 fMRI data analysis	40
3.3 Results	42
3.3.1 Behavioral data	42
3.3.2 fMRI data	44
3.4 Discussion	50
3.4.1 Behavioral results.....	50
3.4.2 Sound azimuth-related BOLD signal changes with the auditory-only condition	52
3.4.3 Effects of visual stimuli on auditory spatial perception...	54
3.4.4 Correlation between the neural responses and the behavioral data	58
3.4.5 Plasticity in human sound localization	59
3.4.6 Ventriloquism effect in relation to the auditory dorsal pathway	60

3.5 Conclusion	61
Learning in communication: Learning-related changes in neural responses to linguistic stimuli	63
Chapter 4. Difference in neural responses to native and non-native linguistic stimuli	64
4.1 Introduction	64
4.2 Materials and methods	67
4.2.1 Subjects	67
4.2.2 Stimuli and procedure.....	68
4.2.3 Functional magnetic resonance imaging data analysis	69
4.3 Results	70
4.3.1 Behavioral performance	70
4.3.2 Brain imaging	70
4.4 Discussion	74
4.5 Conclusion	76
Chapter 5. Neural response changes related to learning of non-native linguistic stimuli	78
5.1 Introduction	78
5.2 Methods	87
5.2.1 Subjects	87
5.2.2 Overall experimental design	88
5.2.3 MRI data acquisition and preprocessing	90
5.2.4 fMRI data analysis	91
5.3 Results	92

5.3.1 Behavioral performance	92
5.3.2 Brain imaging	94
5.4 Discussion	99
5.4.1 Letter-specific visual form processing	100
5.4.2 Grapheme-to-Phonological Conversion Processing	101
5.5 Conclusions	105
Chapter 6. Conclusions	107
6.1 Summary of the dissertation	107
6.2 Future directions	110
REFERENCES	111

Chapter 1

Introduction

1.1 Why we study human-to-human communication with fMRI?

‘Communication’ is a very broad topic. Diverse fields of research have studied various aspects of ‘Communication’ with varying goals in regards to theory and application. ‘Communication’ is also a difficult word to define. According to Merriam-Webster Online (<http://www.merriam-webster.com/>), communication means “the act or process of using words, sounds, signs, or behaviors to express or exchange information or to express your ideas, thoughts, feelings, etc., to someone else”. Thus, a basic communication model consists of four components: a sender, a message, a medium and a recipient. The sender and recipient can be human or non-human. However, in our study, both of them are human as our primary focus on is human-to-human communication. Although there are various media of transmission used for human-to-human communication, most prominent media are human speech and language (Cherry, 1990). A primary goal of our research is to understand the neural mechanisms involved in processing language-based communication in humans.

In the past, our knowledge about how the brain processes human language was primarily obtained by studies of brain-injured patients (i.e. lesion studies). In lesion studies, human brain function is localized by studying the correlation between the site of the brain injury and the behavioral disorder. Among those studies, the most famous are the brain lesion studies of Paul Broca and Karl Wernicke. Broca (1861) investigated patients who lost the ability to speak after

injury to the posterior inferior frontal gyrus of the brain. This type of aphasia is called Broca's aphasia or expressive aphasia and this region of the brain is referred as Broca's area. Broca's area is considered to play a key role in speech production because Broca's aphasiacs typically have impaired speech production and relatively spared speech comprehension (Broca, 1861). On the other hand, (Wernicke, 1874) investigated patients who are unable to understand the language in its written or spoken form after injury to the posterior section of the superior temporal gyrus. This type of aphasia is called Wernicke's aphasia or receptive aphasia and this brain region is referred to as Wernicke's area. Wernicke's area is considered to play a key role in language comprehension because Wernicke's aphasiacs typically have impaired language comprehension and relatively spared speech production even though their speech lacks meaning (Wernicke, 1874).

Lesion studies advanced our understanding of brain function but have several limitations. For example, one of the limitations is related to the network structure of the brain (Rorden & Karnath, 2004). When a certain behavior is impaired by damage to a specific brain region, we can tell that the brain region is necessary for the behavior but cannot tell that it is sufficient for the behavior. How the brain region is involved with a certain cognitive processing cannot be determined by lesion studies alone. Another limitation is based on the difficulty of finding patients with isolated brain damage (Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004). Moreover, the behavioral impairment associated with brain damage often diminishes over time because of healing or compensation by other brain regions (brain plasticity and degeneracy) (Rorden & Karnath, 2004). At present, advancement of neuroscience technology allows us to measure brain activity directly in healthy individuals. With this emergence of brain imaging techniques, classical views of language processing based on lesions data have been

modified (Granacher, 2015). For example, a recent voxel-wise lesion-symptom mapping study suggest that persons with Broca's aphasia have damage to both Broca's and Wernicke's areas (Fridriksson, Fillmore, Guo, & Rorden, 2014). The convergence of knowledge from lesion studies and brain imaging studies gives us new insights into the anatomical foundations of human brain function (Rorden & Karnath, 2004).

There are various techniques to study human brain function. Brain activation is measured by electrophysiological or functional imaging techniques. Electrophysiological techniques, such as electroencephalography (EEG) or magneto-encephalography (MEG), provide high temporal resolution on the order of milliseconds but poor spatial resolution. In contrast, functional imaging techniques, such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) provide high spatial resolution and good temporal resolution (on the order of seconds). Of these techniques, fMRI is the most popular technique to study human brain function at present. This is probably because it is non-invasive (doesn't involve radiation) and because it is relatively easy for the experimenter to collect and analyze data. In this dissertation, we shed new light on human-to-human communications by fMRI experiments.

1.2 About fMRI studies

fMRI measures neural processing carried out in the brain by detecting changes in blood oxygenation and flow that occur in response to neural activity (Logothetis, 2002). The change of the magnetic resonance image contrast reflecting the blood oxygenation level was first reported by Seiji Ogawa and colleagues (Ogawa, Lee, Kay, & Tank, 1990) and is called blood oxygenation level-dependent (BOLD) effect. Unlike PET,

the BOLD fMRI signal has no absolute interpretation. As a result, fMRI experiments test for differences in the magnitude of the signal between different conditions and it is very important to include suitable control conditions or baselines that are only different in the variable of interest (Gusnard, Raichle, & Raichle, 2001). High-level cognitive processing like language produces robust activations not only the language-sensitive regions but also in brain regions also involved with non-linguistic processing (Fedorenko, Behr, & Kanwisher, 2011). Therefore, with regards to localizing brain regions involved with specific functions, it is important to choose experimental paradigms that aim at simple and specific processing of language (e.g. phonological aspect of language).

1.3 Purposes of the study

In human verbal communication, it is important to identify who is talking and to understand what the person is talking about. The former is related to spatial processing and the latter is related to learning. In order to identify who is talking, an individual needs to localize where the talker is located and in order to understand what the person is talking about, an individual needs to learn specific aspects of the language being used. This dissertation consists of two major purposes: spatial processing and learning. In the spatial processing section, we discuss how the brain integrates conflicting audiovisual spatial information. In the learning section, we discuss how learning affects neural responses to linguistic stimuli.

1.3.1 Spatial processing in communication: Integration of audiovisual spatial information in speech perception.

Audition and vision are the two major modalities to perceive speech. Since very early in life, we see the lip movements of speakers and hear their voices at the same time. Auditory and visual information of speech is strongly associated with our daily lives. A strong influence of visual information on speech perception has been demonstrated by the McGurk effect (McGurk & MacDonald, 1976) that demonstrates that visual information from a speaker's face affects auditory speech perception. For example, when you see a person is saying /ga/ and you hear /ba/, you perceive the sound as /da/. Moreover, it has been well studied that the speaker's facial movements contribute to speech intelligibility (for a review, see Massaro, 1987). Therefore, for efficient verbal communication, associating and integrating auditory and visual information of speech are very important. The ability to associate audio-visual information is acquired within the first few months of life (Kuhl & Meltzoff, 1982). Lack of audio-visual integration during a critical period affects the ability to associate audio-visual information of speech. This was demonstrated by impairments in audio-visual integration ability in children with hearing loss who received cochlear implants after 30 months of age (Schorr, Fox, van Wassenhove, & Knudsen, 2005).

It is reasonable to assume that prolonged exposure to concurrent audio-visual information leads to the automatic association of this information in the brain. Even if auditory and visual speech signals are presented from different locations, they are often perceived to originate from the same object or speaker. Under these conditions, the human brain associates and integrates features of auditory and visual information. The ventriloquism effect in which vision (e.g. lip movements) affects the localization of a sound (e.g. a speaker's voice) is a good example of the audio-visual association when their locations are different. Although the phenomena of the ventriloquism effect have

been studied extensively (for a review, see Witten & Knudsen, 2005), its neural basis is still unknown.

In this study, we tried to determine how the brain associates auditory and visual spatial information in speech perception. Unlike the visual system, neural mechanisms for auditory localization are still under debate. Therefore, in Chapter 2, before we investigate the integration of auditory and visual speech information in space, we tried to specify brain regions involved with auditory spatial processing in the real world. Specifically, we investigated differences in the neural responses between realistic spatial (externalized) sounds and artificial (internalized) sounds. Then, in Chapter 3, we investigated the neural basis of audiovisual spatial integration in speech perception by using the ventriloquism effect.

1.3.2 Learning in communication: Neural response differences between native and non-native linguistic rules

Language plays a fundamental role in human-to-human communication. Human language is governed by many rules such as phonological, semantic, syntactic, pragmatic, prosodic and idiosyncratic rules (Roby-Brami, Hermsdorfer, Roy, & Jacobs, 2012). People have to learn those rules to communicate with one another. In order to understand the neural mechanisms of language processing, it is important to study how learning modifies neural representations of linguistic stimuli.

Among various linguistic rules, phonological learning starts at the earliest time in life (Kuhl et al., 2008). In phonology, phonemes are the units in which sounds can be grouped distinctively within a language. Young infants perceive both native and non-native phonemes contrastively but by the end of their first year, they are tuned into the phonemic categories of their language environment (Werker et al.,

2007). Adults, on the other hand, have difficulty discriminating non-native phonemic contrasts that are assimilated to a single phoneme in their native language (Miyawaki, Liberman, Jenkins, & Fujimura, 1975). In Chapter 4, we compared neural correlates of a native phonemic contrast with neural correlates of a difficult non-native phonemic contrast. The purpose of this fMRI experiment is to examine whether different learning levels cause different neural responses in phoneme identification. For this purpose, we used the Japanese durational contrast that is difficult for native English speakers and compared neural responses to the Japanese durational contrasts between native English and native Japanese speakers.

Although comparisons of native and non-native linguistic processing allow us to speculate on whether neural representations of learned linguistic stimuli are different from neural representation of not-learned linguistic stimuli, it does not prove that learning causes modification of neural representation. The purpose of the fMRI experiment in Chapter 5 is to test whether learning can modify neural representations of linguistic stimuli. For this purpose, we compared brain activity before and after learning a non-native letter (phonogram) and investigated whether learning can modify neural representations of non-native orthographic to phonological mappings.

1.4 Outline of the dissertation

This dissertation consists of six chapters. FMRI studies reported in Chapters 2-5 are published in peer-reviewed journals by the author of this dissertation as the first author. Fig. 1.1 is a diagram to show the outline of the dissertation.

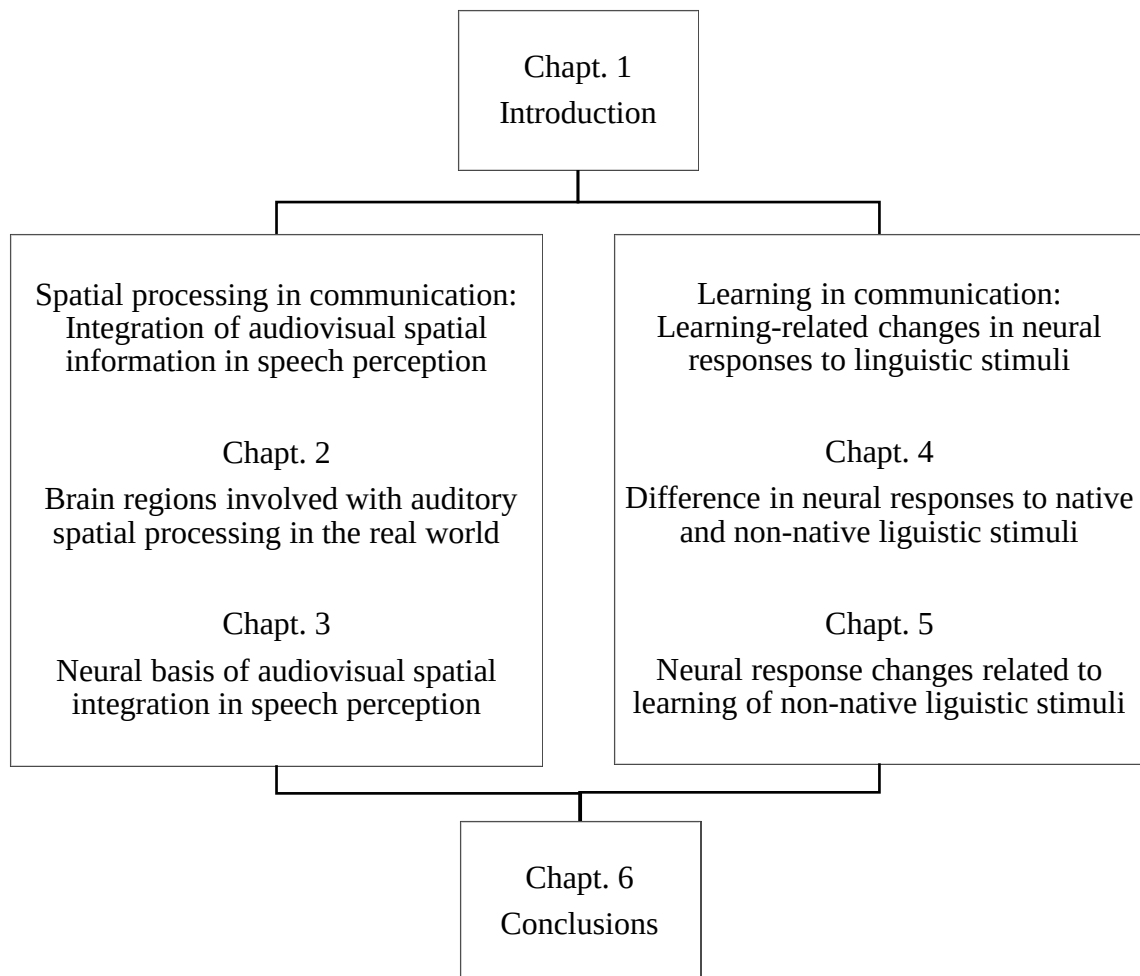


Figure 1.1. Outline of the dissertation.

Brief descriptions of Chapters 2-5 are the following:

Spatial processing in communication: Integration of audiovisual spatial information in speech perception.

- Chapter 2: Brain regions involved with auditory spatial processing in the real world. Differences in neural responses to realistic and artificial spatial sounds were investigated by comparing externalized and internalized auditory stimuli

- Chapter 3: Neural basis of audiovisual spatial integration in speech perception. How visual localization affects neural processes of auditory localization in speech perception was investigated by utilizing the ventriloquism effect.

Learning in communication: Learning-related changes in neural responses to linguistic stimuli.

- Chapter 4: Difference in neural responses to native and non-native linguistic stimuli. Differences in neural responses to native and non-native phonemes were investigated by comparing native English and native Japanese speakers for perception of a Japanese phonemic (durational) contrast that does not occur in English.
- Chapter 5: Neural response changes related to learning of nonnative linguistic stimuli. Changes in neural activity related to learning non-native orthographic to phonological mappings were investigated by comparing brain activity before and after learning previously unknown phonograms.

Spatial processing in communication: Integration of audiovisual spatial information

For efficient human-to-human communication, identifying a speaker's location and integrating auditory and visual cues of speech is very important. In this section, we discuss how the brain associates auditory and visual spatial information in speech perception. We investigated brain regions involved with auditory spatial processing in the real world in Chapter 2 and the neural basis of audiovisual spatial integration in speech perception in Chapter 3.

Chapter 2

Brain regions involved with auditory spatial processing in the real world¹

2.1 Introduction

Prior to investigation of audiovisual spatial integration in speech perception, we examined brain regions involved with auditory spatial processing in the real world. When sounds reach our ears, the acoustic waveform vibrates our eardrums and through a series of steps is transformed into neural impulses and propagated to multiple brain regions. Location of the sound sources are represented in time and level differences between left and right ears (interaural time difference: ITD, interaural level difference: ILD), spectral characteristics arising from the direction-dependent diffraction and reflection properties of the head, pinna, and torso (head-related transfer functions: HRTFs), and echoes of the sounds under reverberant environments (Blauert, 1997). Thus, the brain processes complex auditory signals, and as a result of the processing, we hear the characteristics of sound including pitch and loudness, determine direction and distance of sound-sources, and sense information about the environment around us. While it is important to be able to identify what a sound is, it is also important to be able to detect sound-source location and to predict direction of a moving sound source in order to focus attention to events for survival in many species,

¹ This chapter is published as Callan, Callan, and Ando 2013.

especially when visual cues are not available.

Sound-sources are localized by interaural (i.e. binaural) and spectral cues. Interaural cues, such as ITD and ILD, are primary cues for localization in the horizontal plane (Blauert, 1997). Spectral cues, such as HRTFs, are important for estimation of elevation, localization in the median plane, and resolution of front-back confusion (Blauert, 1997). Moreover, spectral cues are necessary in order to experience the sound as localized outside the head instead of originating inside the head (Hartmann & Wittenberg, 1996; Plenge, 1974). In addition to those cues, we also perceive reverberation of a room in everyday settings. Room reverberation distorts spatial cues for sound localization but is an absolute cue for auditory distance (Bronkhorst & Houtgast, 1999; Mershon & Bowers, 1979). Room reverberation also provides information on room characteristics, and improves the subjective realism and externalization achieved in a headphone simulation (Shinn-Cunningham, Kopco, & Martin, 2005).

Spatial processing in the human auditory cortex has been investigated by functional magnetic resonance imaging (fMRI) studies using either internalized (inside-the-head located) auditory stimuli (e.g., Baumgart, Gaschler-Markefski, Woldorff, Heinze, & Scheich, 1999; Krumbholz et al., 2005) or externalized (outside-the-head located) auditory stimuli (e.g., Deouell, Heller, Malach, D'Esposito, & Knight, 2007; J. D. Warren, Zielinski, Green, Rauschecker, & Griffiths, 2002). Both types of studies suggested the involvement of the posterior superior temporal gyrus (pSTG) in spatial processing. However, no fMRI studies have compared neural activity between internalized and externalized auditory stimuli in one study. Neural responses to natural complex stimuli are very likely to be different from neural responses to artificial simple stimuli. In fact, electroencephalography (EEG) and magnetoencephalography (MEG) studies that compared internalized

stimuli with externalized stimuli report much stronger responses (Getzmann & Lewald, 2010a) and a larger dynamic range of the N1m responses (Palomaki, Tiitinen, Makinen, May, & Alku, 2005) for the externalized stimuli. Furthermore, Palomaki et al. (2005) found that neural responses to individualized HRTF stimuli including room reverberation predicted behavioral performance better than neural responses to non-individualized HRTF stimuli.

How these neurons represent auditory space remains unknown though considerable neuroimaging studies indicate the existence of spatially selective neurons in the pSTG. Auditory space can be separated into left-right, front-back and up-down dimensions. Among them, the left-right dimension has been extensively studied. The auditory cortices receive both contralateral and ipsilateral projections from each ear but the cortical neurons respond more strongly to sounds presented monaurally to the contralateral ear (e.g., Schonwiesner, Krumbholz, Rubsamen, Fink, & von Cramon, 2007; Woldorff et al., 1999). Thus, contralateral representation of auditory space in the pSTG is expected but it is not clear whether binaurally presented spatial sounds are processed contralaterally. Several neuroimaging studies (Krumbholz et al., 2005; Palomaki et al., 2005; Salminen, Tiitinen, Miettinen, Alku, & May, 2010) have reported contralateral responses to binaurally presented stimuli but, on the other hand, several other studies (Brunetti et al., 2005; Woldorff et al., 1999; Zimmer, Lewald, Erb, & Karnath, 2006; Zimmer & Macaluso, 2005) have reported no contralateralized processing. The type of auditory stimulus (i.e. externalized realistic spatial sounds vs. internalized spatially impoverished sounds), the state of the auditory stimulus (i.e. stationary vs. moving), and sensitivity of the measurements (i.e. N1m amplitudes vs. BOLD responses) were considered to be reasons for these contradictory results. A few neuroimaging studies (Getzmann & Lewald,

2010b; Pavani, Macaluso, Warren, Driver, & Griffiths, 2002) comparing horizontal and vertical sound movements suggest that azimuth and elevation components of dynamic auditory spatial information are processed in common cortical substrates. As for front-back dimension, as far as we know, no neuroimaging has investigated the cortical representation.

The purpose of this study is to investigate differences in how externalized sounds and internalized sounds are processed in the human brain. In psychoacoustics, outside-the-head localization is referred to as "localization" and inside-the-head localization is referred to as "lateralization". Therefore, in other words, we are investigating differences in neural processing of localization and lateralization. For this purpose, we compared three types of sound recording methods; binaural, stereo and mono. All three types of recordings were done in the same reverberant room so that all auditory stimuli included the same room reverberation. Both binaural and stereo recording methods used two microphones so that they both include ITD and ILD cues. The mono recording method that used one microphone did not include ITD and ILD cues. In the binaural recording, microphones were positioned at the entrance of the participant's ear canals so that the recorded sounds included individualized HRTFs that allow participants to experience externalized auditory space using headphones. Based on previous EEG and MEG studies (Getzmann & Lewald, 2010a; Palomaki et al., 2005), we hypothesized that externalized auditory stimuli activate the cortical substrate involved in auditory spatial processing more than internalized auditory stimuli. In addition, we investigated neural correlates of auditory localization in the left-right and front-back dimensions. Because the low sensitivity of functional imaging methods was considered as one of the reasons for the lack of contralaterality in several previous studies (Werner-Reiss & Groh,

2008), we performed a region of interest analysis in the bilateral STG for this investigation. The elevation (up-down dimension) of the sound source is computed predominantly from monaural spectral cues and is assumed to be processed in the pSTG (Getzmann & Lewald, 2010b; Pavani et al., 2002). Since front-back location is also computed from monaural spectral cues, we expected modulation of brain activity in the pSTG for front/back localization as well.

2.2 Materials and Methods

2.2.1 Participants

Fourteen adults (7 male; 22-40 years of age, mean 27.5) participated in Experiment 1. About one month later, twelve of them (6 male, 22-40 years of age, mean 28.3) participated in Experiment 2. All participants had no neurological or psychiatric history, had pure-tone thresholds within normal range (≤ 20 dB HL) for octave frequencies between 250 and 8000 Hz. Thresholds were obtained with 5 dB steps and levels lower than 15 dB were not tested. Five subjects had 20 dB thresholds at either or both 250 Hz and 500 Hz only in their right ears. Other than that, their thresholds are at or better than 15 dB. We did not check participants' precise interaural threshold differences. The interaural threshold differences may influence localization processing. However, in people with normal hearing, plasticity in spatial hearing has been demonstrated by the ventriloquism aftereffect (e.g., Recanzone, 1998) and unilateral earplugging experiments (e.g., Kumpik, Kacelnik, & King, 2010). Therefore, within normal hearing range, we assume that effects of interaural threshold differences are not significant. Moreover, individual differences were controlled by employing a random effects model for group analysis. All participants gave written informed

consent for experimental procedures approved by the institutional review board at the National Institute of Information and Communications Technology.

2.2.2 Stimuli and procedure

The auditory stimuli were recorded in three ways: binaural, stereo and mono. The sound used for recording was uniformly distributed noise (band-pass filtered white noise; cutoff frequency, 0.6 kHz and 22 kHz; duration 100ms; rise and fall times, 20 ms) (Zimmer & Macaluso, 2005). The noise was presented through a loudspeaker (Eclipse TD508II, Fujitsu Ten Ltd.) from one of 12 horizontal directions (0°, 30°, 60°, 90°, 120°, 150°, 180°, -150°, -120°, -90°, -60°, -30°, angle 0° was in front of the subject, angles > 0° were in the right hemifield) with a distance of 2 m from the participant. In order to avoid variance caused by different acoustical characteristics of different speakers, only one speaker was used.

Recording methods were as follows. For the binaural recording (BR), the participant sat down in a chair. In-ear binaural microphones (SP-TFB-2, The Sound Professionals Inc.) were positioned at the entrance of the participant's ear canals, and stimuli were recorded through the microphones. For the stereo recording (SR), the same microphones were placed in similar positions to where participants' ears were located (the height from the floor was 115 cm and the distance between left and right microphones was 15.6 cm). For the mono recording (MR), the left microphone was placed in the center of the positions used for stereo recording. The recordings took place in a slightly reverberant room. For each direction, the noise was recorded 18 times and 9 of the recordings were used for the fMRI experiment. Altogether, 108 binaurally recorded sounds (12 directions x 9 times)

were prepared for each participant. 108 stereo and 108 mono sounds were made once and used for all participants. The duration of each of the sound stimuli was 250 ms. In order to include all reverberation, we kept a 150 ms interval in addition to the 100 ms duration of the sound source. Sound levels for each recording type were matched using coefficients that were calculated to match the root mean square energy of the sounds in front of the subjects (angle 0°).

For each trial, participants heard trains of three sounds (sound onset asynchrony 300 ms; duration of each train 850 ms; inter-train interval 2150 ms) and perceived auditory apparent motion (about 18.8 km/h). We divided 12 directions into 4 quadrants. Three sounds that belonged to the same quadrant were presented as a train either in ascending or in descending order (e.g., 0°, 30°, 60° or 60°, 30°, 0°). Auditory stimuli were delivered via MR-compatible headphones (Hitachi Advanced Systems' ceramic transducer headphones; frequency range 30-40,000 Hz, approximately 20 dB SPL passive attenuation). There were three experimental conditions defined by the recording method (binaural, stereo, mono) as well as a null condition in which the subject did nothing. We employed a rapid event-related paradigm and the conditions were presented randomly.

In Experiment 1, participants held a 2-button fiber optic response pad (Current Designs Inc.) in each hand. They were instructed to respond to which quadrant a sound source was located in by pressing one of four buttons (e.g., 0°, 30°, 60° → right hand's front button, -120°, -150°, -180° → left hand's back button) and to press any button twice if they could not localize the stimuli at all. Therefore, chance level for the task was 20 %. In Experiment 2, participants held a 4-button diamond fiber optic response pad (Current Designs Inc.) in their right hands. They were instructed to respond to identify the type of stimulus (BR, SR or MR) by pressing one of three buttons and chance level was

33 %. We explained to subjects the characteristics of the different stimuli: 1. Mono stimuli do not have any localization cues, 2. Stereo stimuli have left-right cues but do not have front-back cues and are localized inside our heads, and 3. Binaural stimuli have both left-right and front-back cues and can be localized outside our heads. Both experiments consisted of three runs that each contained 16 trials per condition. The duration of each run was approximately 8.5 minutes. Experiment 1 and 2 were identical except for the tasks.

2.2.3 MRI data acquisition and preprocessing

For structural and functional brain imaging, a 3T scanner (Siemens MAGNETOM Trio, A Tim system) was used at the ATR Brain Activity Imaging Center. Functional T2*-weighted images were acquired using a gradient echo planar imaging sequence (TR = 3000 ms, matrix size = 64 x 64 pixels, field of view = 192 x 192 mm, 30 slices). The acquisition time (TA) was 1800 ms so there was a 1200 ms quiet period between scans. We presented auditory stimuli during these inter-scan intervals so that the presentation of auditory stimuli was not interrupted by MRI scanning noise. The slice thickness was 5 mm and 170 volumes per run were obtained. The first two scans from each run were discarded to allow for T1 equilibration effects.

Images were preprocessed using programs within SPM8 (Wellcome Department of Cognitive Neurology, London). Images were realigned and spatially normalized using a template defined by the Montreal Neurological Institute (MNI), and they were smoothed using a twice voxel size ($6 \times 6 \times 10$ mm) FWHM Gaussian kernel. Before the acquisition of functional images, T2-weighted anatomical images were acquired in the same plane as the functional images (matrix size = 256×256 pixels). The T2-weighted images were co-registered with the mean

of the functional images and used to calculate the parameters for spatial normalization of functional images.

2.2.4 fMRI data analysis

For both experiments, preprocessed MRI data were analyzed statistically on a voxel-by-voxel basis using SPM8 (128 s high pass filter, serial correlations corrected by an autoregressive AR (1) model). The task-related neural activity was modeled with a series of events convolved with a canonical hemodynamic response function (HRF). Six movement parameters derived from the realignment were also included in the model. Contrasts against baseline were calculated for each condition, and participant specific contrast images were used as inputs for the second level analysis (i.e. random effects model analysis). We conducted a within-subjects one-way ANOVA for overall group analysis. In Experiment 2, we employed bilateral superior temporal gyrus templates of the automated anatomical labeling (AAL) for the small volume correction in SPM and further analysis were performed for the left vs. right and the front vs. back sound source processing. The same comparisons using the small volume correction were not performed in Experiment 1 because neural responses to the button press response task severely confounded neural responses to sound locations. In both experiments, height and extent thresholds were $p < 0.05$, corrected for family wise error.

2.3 Results

2.3.1 Behavioral data for experiment 1

For each quadrant of each recording condition, percentage of correct responses were calculated for each participant. A within-subjects two-way ANOVA shows the significant differences in the main effect of recording conditions ($F[2,26] = 126.89$, $p < 0.01$) and in the interaction ($F[6,78] = 3.54$, $p < 0.01$). Participants could localize the BR stimuli better than the SR stimuli. On the BR condition, a simple comparison of right vs. left direction revealed that left sound sources were localized more accurately than right sound sources ($F[1,13] = 9.30$, $p < 0.01$) but a significant difference was not found for the front-back comparison. The left-hemifield dominance was not found for the SR nor MR condition.

2.3.2 fMRI data for experiment 1

In order to investigate differences between externalized and internalized sound processing, we compared the BR and SR conditions. The BR-minus-SR contrast showed enhanced neural activity in the bilateral posterior superior temporal gyri (pSTG) including planum temporale (PT), posterior middle temporal gyri (pMTG), superior and inferior parietal lobules, right caudate (Table 2.1 and Fig. 2.1 center). Mean beta values (parameter estimates) at peak voxels in the pSTG are also plotted in Fig. 2.1. Blood oxygen level dependent (BOLD) signal in the right pSTG was larger during stimulation in the contralateral (left) than in the ipsilateral (right) hemifield. Likewise, the BOLD signal in the left pSTG was larger during stimulation in the contralateral (right, especially right-back) than in the ipsilateral (left) hemifield. No

significant differences were found for the opposite (SR-minus-BR) contrast.

2.3.3 Behavioral data for Experiment 2

Percentage of correct responses for each recording condition were calculated. A within-subjects one-way ANOVA did not show any significant differences between the three recording conditions (BR = 55.7 %, SR = 50.6 %, MR = 67.5 %, $F[2,22] = 3.054$, $p = 0.07$). On average, participants could identify recording types significantly better ($p < 0.01$) than chance level (33%). For each subject, we plotted percentages of the BR and SR responses for BR and SR stimuli separately (Fig. 2.2). As the figure shows, individual differences were quite large for perceiving those stimuli. Four participants responded almost identically to both BR and SR stimuli indicating that they could not differentiate those stimuli.

2.3.4 fMRI data for experiment 2

We compared the BR and SR conditions. The BR-minus-SR contrast showed significant differences in the bilateral pSTG including the PT and left pMTG (Table 2.1 and Fig. 2.1 center). Consistent with the results from Experiment 1, contralateral BOLD responses were observed in bilateral pSTG. However, overall mean beta values of the left pSTG were smaller in Experiment 2 than Experiment 1 (Fig. 2.1 sides). Region of interest analysis in bilateral superior temporal gyri were performed in order to test for a significant difference between the left and right sound sources and a significant difference between the front and back sound sources. For the BR condition, significant contralateral activation was found in voxels in bilateral pSTG (Fig. 2.3). Moreover,

significant activation for the back relative to the front sound sources were found in the left pSTG (Fig. 2.3). For the SR condition, significant contralateral activity was only found in the right pSTG and the front-back comparisons did not show any significant differences. Neither difference was found for the MR condition.

Table 2.1: fMRI findings for the contrast of the externalized > internalized sounds¹

Region		BA	Cluster	MNI coordinates			z
			size	x	y	z	value
<i>(a) Experiment 1</i>							
Superior temporal gyrus	R	22	128	62	-34	16	6.45
	L	22	104	-62	-32	13	6.45
Middle temporal gyrus	L	37	181	-52	-58	-2	6.83
	R	37	195	52	-52	1	6.23
Caudate	R		80	16	2	10	5.56
Superior parietal lobule	R	40	58	34	-56	55	5.47
Inferior parietal lobule	L	7	42	-32	-48	55	5.25
<i>(b) Experiment 2</i>							
Superior temporal gyrus	R	22	115	62	-32	13	5.88
	L		81	-58	-36	22	5.48
Middle temporal gyrus	L	21	140	-48	-54	13	5.45

¹ Callan, Callan, and Ando 2013:24, table 1.

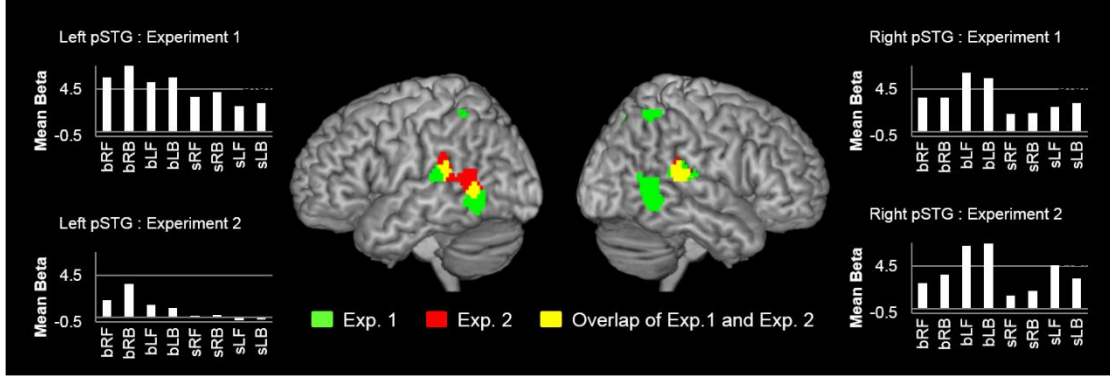


Figure 2.1. Brain areas activated by the externalized-minus-internalized auditory stimuli (i.e. $BR > SR$)¹. At the center of the figure, brain activation patterns from two experiments were rendered on the surface of the brain. Green corresponds to Experiment 1 results, red corresponds to Experiment 2 results and yellow corresponds to overlap of two results. On both sides, mean beta values (a.u.) at the peak voxel in the pSTG are plotted for each stimulus type. For each stimulus type, the first lower case character corresponds to a recording type (b, binaural; s, stereo) and the following upper case characters correspond to a location of sound sources (RF, right front; RB, right back; LB, left back; LF, left front).

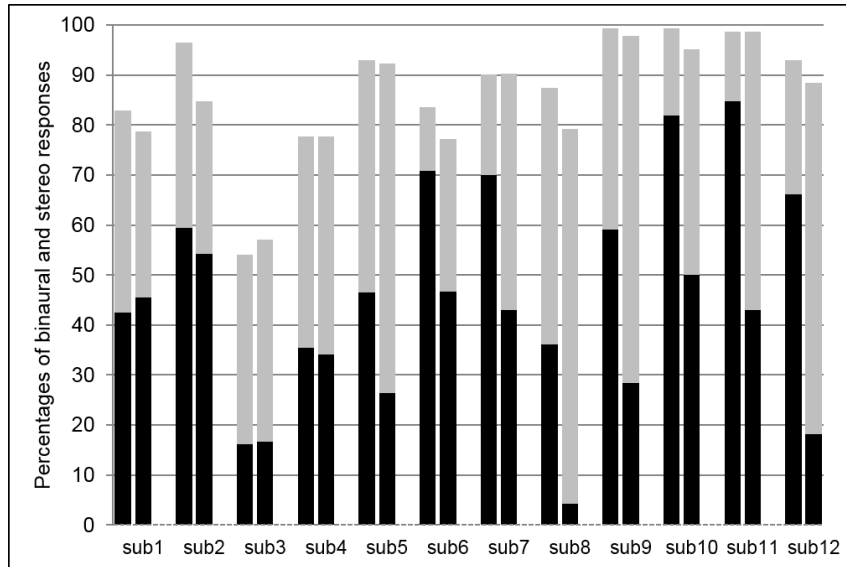


Figure 2.2: Behavioral data for Experiment 2 on the stimulus type identification task². There are two bars for each subject. The first bar corresponds to the binaural recording stimuli and the second bar corresponds to the stereo recording stimuli. For each bar, the black part indicates a percentage of binaural responses and the grey part indicates a percentage of stereo responses. The chance level is 33 %.

¹ Callan, Callan, and Ando 2013:25, figure 1.

² Callan, Callan, and Ando 2013:25, figure 2.

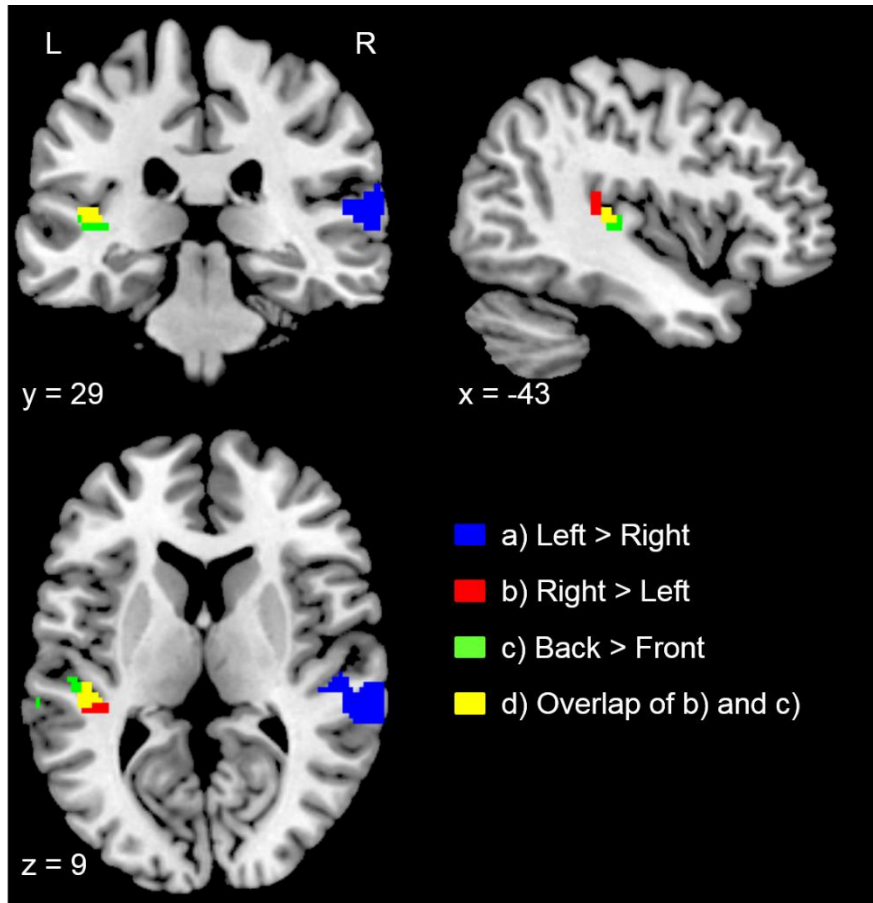


Figure 2.3: Summary of region of interest analysis in the pSTG¹. Brain activations were overlaid on a ch2better.nii.gz template using MRICron. For this figure, a threshold of $p < 0.001$ was used. Blue corresponds to the left-minus-right contrast. Red corresponds to the right-minus-left contrast. Green corresponds to the back-minus-front contrast. Yellow corresponds to the overlap of the right-minus-left and the back-minus-left contrasts.

2.4 Discussion

This is the first fMRI study to investigate neural activation differences between localization (outside-the-head localization) and lateralization (inside-the-head localization) in one study. In this study, we compared three recording methods (binaural, stereo and mono recordings). Our results indicate that the externalized realistic spatial sounds, compared to the internalized ones, enhance neural responses in the bilateral pSTG

¹ Callan, Callan, and Ando 2013:26, figure 3.

involved in auditory spatial processing. Using region of interest analysis, we compared differences between left and right sound source locations and differences between front and back sound source locations. Both left and right pSTG were activated more by contralateral auditory stimuli. Moreover, we found that voxels in the left pSTG were activated more by back locations than front locations.

In our first experiment with the active spatial localization task, the behavioral data showed that participants could localize BR stimuli best and that sounds in the left hemifield were localized better than sounds in the right hemifield. Considering that front/back localization cues are only available for the BR stimuli, better performance for the BR stimuli was to be expected. More accurate performance for the left sound sources on the BR condition was consistent with a previous behavioral study (Burke, Letsos, & Butler, 1994) supporting right hemispheric dominance in auditory localization.

In our second experiment using the sound-type identification task, the behavioral results showed that, on average, participants could classify the three recording methods. However, if we look at each individual's data, we found that four out of the twelve participants responded to the binaural and SR stimuli in the same manner (Fig. 2.2), indicating that those participants could not differentiate between BR and SR stimuli. Because we did not use individualized headphone transfer functions, the BR stimuli may fail to produce high fidelity virtual auditory space (Pralong, 1996). It should be noted that even with the headphone transfer function, front-back confusions in the headphone condition have been observed (Wightman & Kistler, 1989) and there might be large individual differences on perception of virtual auditory space. Moreover, we need to mention that the success or failure in differentiation between BR and SR stimuli does not directly address existence or lack of the externalized percept. It is very

difficult to study the externalization of sound images because the externalized percept is subjective and not precisely defined (Hartmann and Wittenberg, 1996). In fact, there is only a limited psychophysical literature on sound externalization. (Blauert, 1997) considered that inside-the-head locatedness occurs when the distance of the auditory event is smaller than the radius of the head and asked participants to report the distance of the auditory event. Plenge (1974) used an experimental apparatus to present internalized and externalized sounds at the same time and asked participants where sound sources were located (intracranial, verged cranial, or extracranial). Hartmann and Witten and Knudsen (2005) synthesized the headphone signals to best resemble real-world sounds so that they were not distinguishable and studied sound externalization, using both discrimination and listener rating, by systematically modifying the synthesis.

Consistent with our hypothesis based on previous EEG and MEG studies (Getzmann & Lewald, 2010b; Palomaki et al., 2005), we found that the externalized auditory stimuli activated the cortical substrate involved in auditory spatial processing (i.e. pSTG) more than the internalized ones (Fig. 2.1). The pSTG activation was found for both experiments with different tasks, indicating the result was reliable. The bilateral regions include the PT. Because activation in the PT has been reported for the analysis of many types of complex sound (e.g. speech, spatial sounds and music) (Altmann, Bledowski, Wibrall, & Kaiser, 2007; A. M. Callan, Callan, Tajima, & Akahane-Yamada, 2006; Price, 2012), Griffiths and Warren (2002) propose the PT as a computational engine for the segregation and matching of spectrotemporal patterns. We assume that the PT was activated more for the externalized sounds to analyze spectral information in the HRTFs. Since shadow effects of the head in binaural recordings enhance the ILD, one may suspect that the neural activity differences observed in this study were due to

different levels of interaural intensity. However, a previous MEG study (Palomaki et al., 2005) reports that binaural recordings elicited a larger dynamic range of the N1m amplitude in the right pSTG than the ILD stimuli that matched the level differences with the binaural recordings and the non-individualized HRTF stimuli. Therefore, it is very unlikely that neural differences in the pSTG in this study were caused solely by different levels of the ILD.

In addition to the pSTG, both experiments showed significant differential activity in the pMTG (Fig. 2.1). Enhanced activation of the pMTG has been observed in several auditory localization studies (Bushara et al., 1999; Seifritz et al., 2002; Zimmer et al., 2006). However, its functional role in auditory localization has not been discussed. In general, the function of the MTG is not well-defined compared to the superior and inferior temporal gyri. Activation in the pMTG has been reported for recognizing environmental sounds (Lewis et al., 2004), perceiving manipulable object motions (Beauchamp, Lee, Haxby, & Martin, 2002), and integrating auditory and visual information about objects (Beauchamp, Lee, Argall, & Martin, 2004). These findings suggest that the pMTG is involved in processing complex stimuli. Differential activity in the parietal lobules and caudate was only found in Experiment 1 (the auditory localization task). The involvement of parietal cortex in coding of auditory space in a coordinate system suitable for movement preparation is well established (e.g., J. D. Warren et al., 2002; Zatorre, Bouffard, Ahad, & Belin, 2002). Animal studies report the involvement of caudate in performance of the conditioned reactions connected with sound localization (Suvorov & Gulyakov, 1988). We assume that differential activity in the parietal lobules was not found in Experiment 2 because the stimulus type identification task did not require to relate sound source locations to the button press responses.

Region of interest analysis with the BR stimuli indicated contralateral representation of auditory space in bilateral pSTG. With the SR stimuli, only the right pSTG showed contralateral activity. Lack of finding contralateral activity for the left pSTG may be caused by smaller contra-ipsilateral modulation elicited by the impoverished spatial sounds compared to the realistic spatial sounds (Palomaki et al., 2005)). Our results confirm that both left and right pSTG are involved in auditory spatial processing and that neurons in the pSTG are more sensitive to the contralateral than ipsilateral sound source locations. Although both right and left pSTG are involved in auditory localization, consistent with many previous studies (e.g., Kaiser, Lutzenberger, Preissl, Ackermann, & Birbaumer, 2000; Palomaki et al., 2005; Zatorre, Bouffard, et al., 2002), the right-hemispheric dominance in auditory localization was inferred from more spread activation (Table 2.1) and larger beta values (Fig. 2.1 sides). This theory was supported by the behavioral result in this experiment in which it was found that participants could localize sound sources located in the left hemifield better than the right hemifield.

Significant activation for the back sounds compared to the front sounds was observed in the left pSTG. Mean beta values at the peak voxels of the left pSTG in both experiments (Fig. 2.1 sides) also indicate that the left pSTG is more sensitive to back than front locations. MEG studies propose that a population rate code of spatial location formed by two opponent populations (i.e. ipsi- and contra-laterally tuned neurons) accounts for sound source lateralization in the horizontal plane (Salminen, May, Alku, & Tiitinen, 2009; Salminen et al., 2010). However, the population rate code cannot account for the front/back discrimination. There is no neurophysiological (EEG or MEG) data to dissociate front/back locations but a recent study reports that a computational model based

on spike timing successfully discriminates front/back locations (Goodman & Brette, 2010). In our study, using externalized auditory stimuli, the left pSTG was found to be activated more for sound sources located in the back but how front/back locations are coded in our brain is still unknown.

2.5 Conclusion

Compared to impoverished spatial auditory stimuli, realistic spatial auditory stimuli enhance neural responses in the pSTG. This may be why we could observe contralateral hemifield preference in bilateral pSTG that many previous research has failed to observe. Both left and right pSTG are involved in auditory localization, but right hemisphere dominance is inferred by both behavioral and neuroimaging data. In addition to contralateral responses, back/front dissociation was observed in the left pSTG. The results suggest that externalized auditory stimuli (i.e. binaural recordings) are more suitable than internalized auditory stimuli (i.e. stereo recordings) to understand auditory neural processing in the real world and that the pSTG is involved with auditory spatial processing in the real world. In the next chapter, we study how audiovisual spatial information is integrated in the pSTG.

Chapter 3

Neural basis of audiovisual spatial integration in speech perception¹

3.1 Introduction

In this chapter, we investigated the neural basis of audiovisual spatial integration in speech perception. In spatial perception, visual information has higher acuity than auditory information and we often misperceive sound source locations when spatially disparate visual stimuli are presented simultaneously. Ventriloquists make good use of this auditory illusion. In this study, we investigated neural substrates of the ventriloquism effect to understand the neural mechanism of multimodal integration.

In ventriloquism, the audience perceives speech sounds as coming from a direction other than their true direction. The speech sounds are localized at the location of the puppet's obviously moving mouth instead of the speaker's unmoving mouth. This ventriloquism effect is one example of 'visual capture' (D. H. Warren, Welch, & McCarthy, 1981). In general, vision dominates (or captures) perception when spatially disparate visual and auditory stimuli are simultaneously presented. In the visual system, the spatial features of the stimulus are encoded by the location of stimulation on the retina. Location information is inherent at the peripheral level. In contrast, spatial

¹ This chapter is published as Callan, Callan, and Ando 2015.

perception in audition is much more challenging because the spatial features have to be extracted from sensory receptors organized according to sound frequency rather than spatial configuration and because auditory cues are easily distorted by echoes and reverberation in the environment. In humans, the spatial resolution of the visual system is superior to the spatial resolution of the auditory system. The ventriloquism effect can be explained as a phenomenon in which the sensory modality with the higher acuity dominates over and captures the other sensory modality with lower acuity (Alais & Burr, 2004; D. H. Warren et al., 1981). Although the phenomena of the ventriloquism effect have been studied extensively (for a review see, Witten & Knudsen, 2005), its neural basis is still unresolved.

To investigate the neural basis of the ventriloquism effect, it is necessary to understand how source location is represented in the human auditory cortex. A topographical place code and a population rate code are two major hypotheses for the coding of auditory localization. The topographical place code assumes that auditory space is represented by the activation of neurons that correspond to particular locations in space (Jeffress, 1948). Conversely, the population rate code assumes that sound source locations are represented by patterns of activity across populations of broadly tuned neurons (Middlebrooks, Xu, Eddins, & Green, 1998). Spatially tuned neurons have been observed in mammalian superior colliculus (Middlebrooks & Knudsen, 1984; Palmer & King, 1982) and in avian inferior colliculus (Knudsen & Konishi, 1978; Takahashi, Moiseff, & Konishi, 1984). However, neurophysiological studies have reported broadly tuned but not narrowly tuned neurons in the auditory cortex in cats (Brugge, Reale, & Hind, 1996; Middlebrooks et al., 1998; Rajan, Aitkin, Irvine, & McKay, 1990) and in monkeys (Werner-Reiss & Groh, 2008). Thus, in mammals, most location-sensitive auditory cortical neurons are broadly tuned and

respond to stimuli located throughout contralateral space. These results suggest that the population rate code (not the topographical place code) is instantiated in the human auditory cortex.

Recent electrophysiological (EEG and MEG) studies using an adaptation paradigm have provided support for the population rate coding of sound azimuth localization in the human auditory cortex (Magezi & Krumbholz, 2010; Salminen et al., 2009). The EEG experiment showed that the adaptation effect was stronger when an interaural time difference changed toward the midline than when it changed away from the midline and that the majority of spatially sensitive neurons in each hemisphere are tuned to the contralateral hemifield (Magezi & Krumbholz, 2010). The MEG experiment showed that all adaptors were effective when the probe and the adaptor were within the same hemifield but not effective when the adaptor was at the midline or in the opposite hemifield (Salminen et al., 2009). Both results indicate that the human auditory cortex represents sound source locations with two populations of broadly tuned neurons. One population is sensitive to the contralateral hemifield and the other population is sensitive to the ipsilateral hemifield, especially to sounds located at lateral extremes (far to the left or to the right side of the perceiver). Thus, the existence of population rate coding in the human auditory cortex is supported by changes in the response level of N1 (the negative deflection at around 100 ms after sound onset). However, which part of the auditory cortex is responsible for the N1 changes and how neural activity levels represent the sound azimuth are still unknown. Compared with EEG and MEG, functional magnetic resonance imaging (fMRI) has higher spatial resolution and is expected to be able to clarify which cortical regions are associated with the population rate code. Nevertheless, there are no fMRI studies that have reported the relationship between the sound azimuth and neural activity. A

single-unit recording study in monkeys reported a monotonic azimuthal sensitivity as well as a contralateral sensitivity at the population level (Werner-Reiss & Groh, 2008). Despite these findings, many fMRI studies fail to show contralateral preference in the human auditory cortex (for a review see, Werner-Reiss & Groh, 2008). Lack of sensitivity of the fMRI measurements as well as the impoverished nature of the stimuli used might be one of the reasons why many studies have failed to find a contralateral preference in the human auditory cortex. To investigate neural mechanisms for auditory spatial perception, we need to enhance neural response levels by presenting optimal stimuli and increase the sensitivity of the fMRI measurements.

Natural complex auditory stimuli that include both interaural and spectral temporal cues (i.e. head-related transfer functions) are considered to elicit more salient neural responses in the posterior superior temporal gyrus (pSTG) than artificial simple auditory stimuli that only include interaural cues (A. Callan et al., 2013; Getzmann & Lewald, 2010a; Palomaki et al., 2005). Using realistic (externalized) auditory spatial stimuli and applying region of interest (ROI) analyses, our previous fMRI study was able to show that both left and right pSTG were more sensitive to sound sources in contra- than in ipsilateral hemifields (A. Callan et al., 2013). As for stimulus content, we chose speech stimuli that are more complex and natural rather than the simple tones and lights used in most previous multimodal integration studies (Vander Wyk et al., 2010). Physiological investigation revealed that most neurons in the auditory cortex respond maximally to sounds located far to the left or to the right but change most abruptly across the frontal midline (Stecker, Harrington, & Middlebrooks, 2005). Stecker et al. (2005) proposed that in rate coding the optimal spatial acuity across the frontal midline observed in behavioral studies is achieved by using the slopes of activity rather than the peaks of activity.

Compared with sounds located far to the left or to the right, sounds located around the frontal midline can induce more abrupt neural response changes. Therefore, if we compare auditory stimuli located around the frontal midline, we may be able to detect a blood oxygenation level-dependent (BOLD) signal change that is monotonically related to the sound azimuth even with fMRI.

In this fMRI study, we first determined regions involved in sound azimuth representation and then investigated how conflicting visual stimuli affect neural activity in these regions. To enhance the detectability of BOLD signal changes, we used the following: (1) realistic spatial sounds that include both interaural and spectrotemporal cues; (2) complex natural speech as stimuli; (3) sound locations around the frontal midline where neural responses show greatest modulation; and (4) ROI analyses. We hypothesized that the regions associated with sound azimuth presentation in the auditory cortex (i.e. pSTG) will show increased BOLD signals monotonically as the sound source changes from the frontal midline to more contralateral locations. We further hypothesized that the ventriloquism effect results from conflicting visual information suppressing the monotonic signal increase associated with sound azimuth representation.

3.2 Materials and Methods

3.2.1 Participants

Sixteen adults (8 male; 22–44 years of age, mean 28.25) participated in this experiment. All participants had no neurological or psychiatric history, had pure-tone thresholds within the normal range (≤ 20 dB HL) for octave frequencies between 250 and 8000 Hz, and gave written informed consent for experimental procedures approved by the

institutional review board at the National Institute of Information and Communications Technology.

3.2.2 Stimuli

The stimuli were audio and video recordings of a female native Japanese speaker articulating Japanese greeting words at a natural speech rate. Five Japanese greeting words were used: "ohayou gozaimasu" (good morning), "konnichiwa" (good afternoon), "konbanwa" (good evening), "arigatou" (thank you), and "sayonara" (good bye). Each of the greeting words were repeated five times. In each video recording the speaker's face, shoulders, and neck were visible (see Fig. 3.1 for a sample frame). Video recordings were made in an anechoic chamber using a digital video camera (30 images/s frame rate, 48 kHz in 16 bits audio sample rate; IVISHV10, JVC Corp., Japan) using a miniature condenser microphone (20 - 20,000 Hz frequency response; DPA4060) attached to the person's collar as audio input. The videos were edited using software (Edius version 4; Canopus) and the duration of each video clip ranged from 0.9 to 1.2 s ("ohayou gozaimasu" 1.2 s, "konnichiwa" 1 s, "konbanwa" 1 s, "arigatou" 0.9 s and "sayonara" 0.9 s). Then, we extracted the speech sounds from the videos. The extracted speech sounds were used for auditory stimuli recordings.

For the recordings, we used binaural and stereo recording methods. The binaural recorded sounds can elicit externalized (outside-the-head) perceptions of auditory space through headphones. In contrast, the stereo-recorded sounds can provide left or right information about the sound sources but cannot elicit externalized perception; therefore, the sounds are localized inside but not outside of the head. For the binaural recording (BR), participants sat down on a chair. In-ear binaural

microphones (20 - 20,000 Hz frequency response; SP-TFB-2; The Sound Professionals Inc.) were positioned at the entrance of the participant's ear canals, and stimuli were recorded through the microphones. For the stereo recording (SR), the same microphones were placed in similar positions to where participants' ears were located (the height from the floor was 115 cm and the distance between left and right microphones was 15.6 cm). The speech sounds were presented through a loudspeaker (55 - 20,000 Hz frequency response; Eclipse TD508II; Fujitsu Ten Ltd.) from one of seven horizontal directions (-30° , -20° , -10° , 0° , 10° , 20° , 30° , left to right, angle 0° was in front of the participant) with a distance of 180 cm from the participants (Fig. 3.1a) and digitally recorded with a 16-bit resolution at a sampling rate of 48 kHz. In order to avoid variance caused by different acoustical characteristics of different speakers, only one speaker was used. For each direction, each greeting word was presented five times. BR stimuli were prepared for each participant. On the other hand, SR stimuli were prepared only once and the same SR stimuli were used for all participants. All auditory stimuli included 5.3 ms no-speech-sound-periods in the beginnings that reflect the distance between the speaker and the listener (180 cm) given the speed of sound (340 m/s). The durations of auditory stimuli were matched with the corresponding video stimuli. Sound levels for each recording type were matched using coefficients that were calculated to match the root mean square (RMS) energy of the sounds in front of the participants (angle 0°).

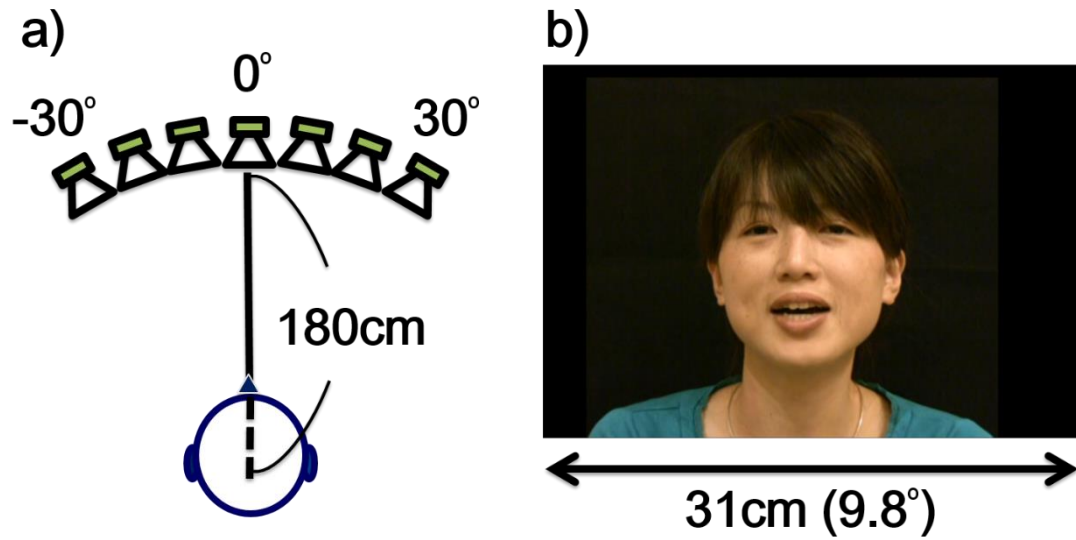


Figure 3.1. Experimental set-up¹. (a) Recording set-up of auditory stimuli. The speech sounds were presented through a loudspeaker from 1 of the 7 horizontal directions (-30° , -20° , -10° , 0° , 10° , 20° , and 30° , left to right, angle 0° was in front of the participant) with a distance of 180 cm from the participants. (b) An example frame of the visual stimuli.

3.2.3 Procedure

For each trial, participants listened to one of the auditory stimuli and responded whether the sound source was located either left or right by pressing one of two buttons. In order to minimize interruption by MRI scanning noise, we presented auditory stimuli during inter-scan intervals. Thus, participants never heard the scanning noise while listening to the stimuli. The left or right forced choice task is the same task used for the ventriloquism aftereffect with macaque monkeys (Woods & Recanzone, 2004). The left-right forced choice task was used instead of localizing perceived stimulus locations for two reasons. The first reason is because the left-right response is considered to be simpler and faster than the localization response so that the task

¹ Callan, Callan, and Ando 2015:3, figure1.

minimizes differences in the hemodynamic response resulting from long response times (Poldrack, 2000). The second reason is because behavioral studies (Jack & Thurlow, 1973; Thurlow & Jack, 1973) of the ventriloquism effect determined that the localization task was not appropriate to measure the ventriloquism effect because the degree of visual capture obtained when subjects are instructed to localize a sound is quite small.

Stimulus presentation and timing was controlled by using Presentation® software (www.neurobs.com). Timing uncertainties were generally smaller than 0.2 ms and the various conditions did not lead to differences in the performance of the program. Auditory stimuli were presented with or without corresponding visual stimuli. When both auditory and visual stimuli were presented, their presentation timings were synchronized. Auditory stimuli were represented in one of seven locations but visual stimuli were always presented in front of the participants. Auditory stimuli were delivered via MR-compatible headphones (Hitachi Advanced Systems' ceramic transducer headphones; frequency range 30 – 40,000 Hz, approximately 20 dB SPL passive attenuation). Mean intensities of auditory stimuli were "ohayou gozaimasu" 66 dB, "konnichiwa" 66 dB, "konbanwa" 67 dB, "arigatou" 65 dB and "sayonara" 67 dB SPL. All visual stimuli were projected on a screen and viewed through a mirror mounted on the MRI head coil at the viewing distance of 180 cm (total display size $9.8^{\circ} \times 7.4^{\circ}$ of visual angle) (Fig. 1b).

We applied a block design paradigm with four experimental conditions (auditory-only BR, auditory-only SR, audiovisual BR, and audiovisual SR) and a baseline rest condition. Each participant performed five 10 min imaging runs. During each run, a 2 min cycle that includes four experimental blocks alternating with a rest block was presented five times. An example of one cycle is “auditory-only BR +

rest + auditory-only SR + rest + audiovisual BR + rest + audiovisual SR + rest". Each experimental block included 7 stimuli (i.e. 7 trials) and lasted 21 s. The rest block lasted 9 s and participants were instructed to simply look at the fixation on the screen. The order of the experimental conditions was randomized for each run. The effect of stimulus length was controlled by presenting each greeting word the same number of times for all conditions. The locations of the auditory stimuli were presented randomly throughout the experiment. In order to control participants' gaze direction, a fixation marker was presented in the center of the screen throughout the experiment except when video stimuli were presented. Participants were instructed to try not to move their eyes and to always look at the center of the screen.

3.2.4 MRI data acquisition and preprocessing

For structural and functional brain imaging, a 3T scanner (Siemens MAGNETOM Trio, A Tim system) was used at the ATR Brain Activity Imaging Center. Functional T2*-weighted images were acquired using a gradient echo planar imaging sequence (TR = 3000 ms, matrix size = 64×64 pixels, field of view = 192×192 mm, slice thickness = 3 mm with a 1 mm gap, 30 slices). For each functional imaging run, 205 volumes were obtained. The first two scans from each run were discarded to allow for T1 equilibration effects. The acquisition time (TA) was 1800 ms so there was a 1200 ms quiet period between scans. We presented auditory stimuli during these inter-scan intervals so that the presentation of auditory stimuli was not interrupted by MRI scanning noise.

Images were preprocessed using programs within SPM8 (Wellcome Department of Cognitive Neurology, London, UK). Images were realigned and spatially normalized using a template defined by the

Montreal Neurological Institute (MNI), and were smoothed using a twice voxel size ($6 \times 6 \times 8$ mm) FWHM Gaussian kernel. Before the acquisition of functional images, T2-weighted anatomical images were acquired in the same plane as the functional images (matrix size = 256×256 pixels). The T2-weighted images were co-registered with the mean of the functional images and used to calculate the parameters for the spatial normalization of the functional images.

3.2.5 fMRI data analysis

For both experiments, preprocessed fMRI data were analyzed statistically on a voxel-by-voxel basis using SPM8 (128 s high pass filter, serial correlations corrected by an autoregressive AR (1) model). The task-related neural activity was modeled with a series of events convolved with a canonical hemodynamic response function. Six movement parameters derived from the realignment were also included in the model. The seven locations (-30° , -20° , -10° , 0° , 10° , 20° , and 30°) of the four presentation types were used as experimental conditions. Left (or right) lateralized activity was tested by a parametric contrast [3, 2, 1, -1.5, -1.5, -1.5, -1.5] (or [-1.5, -1.5, -1.5, -1.5, 1, 2, 3]) for the auditory-only BR condition and auditory-only SR condition separately.

Then, a second-level random-effect analysis was performed to yield statistical parametric maps of the resulting t values for each contrast at the group level. Significant activation clusters were determined using a height threshold of $p < 0.001$ uncorrected and an extent threshold of $p < 0.05$ corrected for family-wise error (FWE). Additionally, small-volume correction (SVC) for multiple comparisons ($p < 0.05$, FWE corrected) was performed using bilateral superior temporal gyrus templates (3875 voxels) in the Automated Anatomical

Labeling atlas (Tzourio-Mazoyer et al., 2002). The use of SVC allows researchers to conduct principled correction using Gaussian Random Field Theory within a predefined region of interest (ROI). To conduct ROI analyses, we used an SPM toolbox called ‘Marsbar’ (<http://marsbar.sourceforge.net/>). ROI analyses were performed in the following steps: (1) ROIs were defined as spheres with a 5-mm radius centered on the peak voxels located in the superior temporal gyrus for the BR and SR conditions separately; (2) we obtained an averaged fMRI signal of all voxels within the spheres and calculated mean beta values for all experimental conditions; (3) for each presentation type, we obtained a slope of left (or right) lateralized activity; (4) significant differences were investigated by comparing between the auditory-only and audiovisual conditions with paired t-tests.

In addition, using behavioral data as regressors, we examined which brain regions were activated more by participants who were more affected by visual stimulus presentation. For this analysis, first, we evaluated laterality of sound source perception by subtracting 50% from percentages of ‘right’ responses at each sound location. Second, for each participant, we averaged absolute values of them and made laterality indexes for the auditory-only and audiovisual conditions separately. Then, we subtracted the audiovisual index from the auditory-only index. Bigger differences indicated stronger ventriloquism effects because the audiovisual average was expected to be smaller than the auditory-only average if the centrally presented visual stimuli captured auditory perception of the laterally presented auditory stimuli. Finally, we tested a linear regression between those behavioral data and BOLD responses (the audiovisual > baseline contrast images) for the BR and SR stimuli separately. For visualization of results, the activation maps are superimposed on a high resolution anatomical MRI brain template (ch2better.nii.gz) using MRICron

software (Rorden, Karnath, & Bonilha, 2007).

3.3 Results

3.3.1 Behavioral data

For each experimental condition, percentages of ‘right’ responses were calculated for each participant and the mean percentages were plotted in the Fig. 3.2. Sound sources located closer to the midline were naturally more difficult to lateralize so that psychometric functions could be generated. In this analysis, 100% ‘right’ responses were expected for the most right stimuli (i.e. 30°), 50% ‘right’ responses were expected for the center stimuli (i.e. 0°), and 0% ‘right’ responses were expected for the most left stimuli (i.e. -30°). Although a response curve with the BR stimuli was consistent with the expected psychometric function, a response curve with the SR stimuli was not. The mean percentage of right responses for 0° was 26.5% instead of 50%. The result indicated that the 0° SR stimuli represented not the center but the left.

In SPSS (IBM SPSS version 22), we performed a within-subjects three-way analysis of variance with presentation type (auditory-only, audiovisual), recording method (BR, SR), and sound source location (-30°, -20°, -10°, 0°, 10°, 20°, 30°) as factors. There were significant main effects of recording method ($F(1,15) = 12.748$, $p < 0.01$) and of location ($F(6,90) = 171.524$, $p < 0.01$). On average, participants perceived the SR stimuli as more left-lateralized than the BR stimuli. Furthermore, the percentage of right responses increased to a greater extent for more right-lateral locations. All three two-way interactions were significant. The first interaction was between the presentation type and the recording method. It indicated that the difference between

the audio-only and audiovisual conditions were bigger for the SR stimuli than for the BR stimuli ($F(1,15) = 5.523$, $p < 0.05$). The second interaction was between the recording method and the sound source location. It indicated that the sigmoid function for the left-to-right locations for the BR stimuli was sharper than the sigmoid function for the SR stimuli ($F(6,90) = 22.189$, $p < 0.01$). The third interaction was between the presentation type and the sound source location. It indicated that the sigmoid function for the auditory-only condition was sharper than the sigmoid function for the audiovisual condition ($F(6,90) = 2.674$, $p < 0.05$). Bonferroni post-hoc tests indicated significant ($p < 0.05$) differences between the auditory-only and audiovisual conditions at -30° , -10° , and 20° . The three-way interaction was not significant.

In addition, as planned comparisons, we performed paired t-tests to investigate in which locations visual stimulus presentation significantly affected auditory perception for the BR and SR stimuli separately. Significant ($p < 0.05$) differences between auditory-only and audiovisual conditions were observed at -30° , -20° , -10° , and 0° with the SR stimuli and -10° and 20° with the BR stimuli (Fig. 3.2). We also tested whether visual stimulus presentation affected left and right sound-source locations differently (i.e. whether differences between auditory-only and audiovisual conditions were larger at left or right sound-source locations). For this analysis, we compared -30° with 30° , -20° with 20° and -10° with 10° for the BR and SR conditions separately. There were no significant differences in these comparisons.

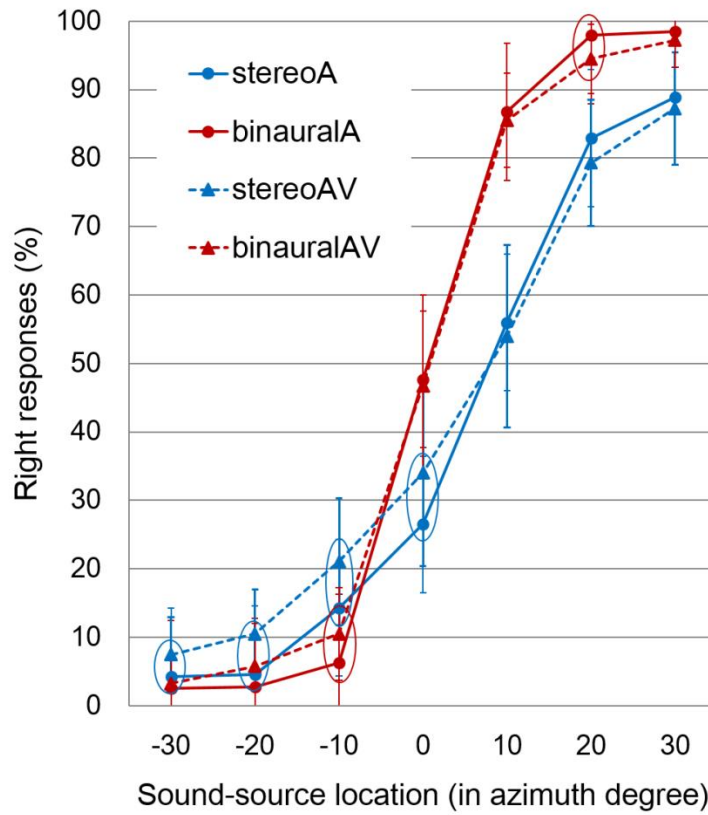


Figure 3.2. Percentage of “right” responses¹. Means and standard errors of the means (SEM) are plotted. On both SR and BR conditions, slopes were steeper for auditory only conditions than for audiovisual conditions. At each sound location, a significant difference ($P < 0.05$) between the auditory-only and audiovisual conditions was indicated by a circle. Stereo A, SR auditory-only; binaural A, BR auditory-only; stereo AV, SR audiovisual; binaural AV, BR audiovisual.

3.3.2 fMRI data

The auditory-only conditions

Under auditory-only conditions, we determined brain regions activated to a greater extent for more lateral sound locations. The results of the whole brain analyses (height threshold of $p < 0.001$ uncorrected and an extent threshold of $p < 0.05$ FWE corrected) for the BR stimuli revealed

¹ Callan, Callan, and Ando 2015:4, figure 2.

that linearly increasing activity for the left lateral sounds existed in the right pSTG (MNI coordinates at the peak voxel $x\ y\ z = 52, -24, 10$, 859 voxels) and in the right precuneus ($x\ y\ z = 8, -50, 42$, 315 voxels), and that linearly increasing activity for the right lateral sounds existed in the left pSTG ($x\ y\ z = -56, -32, 12$, 153 voxels) (Fig. 3.3). Significant increased activity for the contra-lateral sounds in the left and right pSTG were also found by the additional SVC ($p < 0.05$ FWE corrected) using the bilateral superior temporal gyrus templates. The whole brain analyses for the SR stimuli found linearly increasing activity for the right lateral sounds in the left precuneus ($x\ y\ z = -18, -46, 46$, 147 voxels). The whole brain analyses and the SVC failed to show linearly increasing activity in the STG with the SR stimuli.

ROI analyses comparing the auditory-only condition with the audiovisual condition

We performed ROI analyses using peak voxel coordinates in the STG for the BR and SR conditions separately. For the BR condition, the center coordinates for the ROI spheres were $x\ y\ z = -56, -32, 12$ and $x\ y\ z = 52, -24, 10$. Mean beta values in the left and right regions are plotted in Fig. 3.4. A significant decrease in slope for the audiovisual condition compared with the auditory-only condition for the BR stimuli was found in the right pSTG ($t(15) = 1.79$, $p < 0.05$) (Fig. 3.4b). In the right pSTG, visual stimuli presented in the midline reduced neural responses for left lateralized sounds compared to neural responses when those sounds were presented without visual stimuli. For the SR condition, the center coordinates for the ROI spheres were $x\ y\ z = -56, -46, 12$ and $x\ y\ z = 44, -28, 16$. The ROI analyses with the SR stimuli failed to show significant differences between the auditory-only and the audiovisual conditions.

Krumbholz et al. (2005) proposed a hierarchical organization of

auditory spatial processing in which analysis of simple interaural cues begins at the brainstem and more complex signal sensitivity (in their case, moving sound) emerges in the pSTG. From this proposal, it follows that the ventriloquism effect for the simple SR stimuli was not found in the pSTG because the neural change is processed below the level of the primary auditory cortex, in the brainstem. In the brainstem, the superior colliculus (SC) is a site for multisensory integration (Stein & Stanford, 2008). In order to check whether the SC showed linearly increasing activity, we performed additional SVC analyses using spheres with a 5-mm radius centered on the left ($x\ y\ z = -6, -28, -6$) and right ($x\ y\ z = 6, -28, -6$) SC peak voxels reported in a previous study (Linzenbold & Himmelbach, 2012). Linearly increasing activity for the left lateral SR stimuli was not found but linearly increasing activity for the right lateral SR stimuli was found in the left SC ($x\ y\ z = -10, -26, -8$, $p < 0.05$ FWE corrected) (Fig. 3.5). Using contrast estimates of the peak voxel for each participant, we compared the slope of the auditory-only condition with the slope of the audiovisual condition. In contrast to findings in the right pSTG, a significant difference ($p < 0.05$) was found for the SR stimuli but not for the BR stimuli.

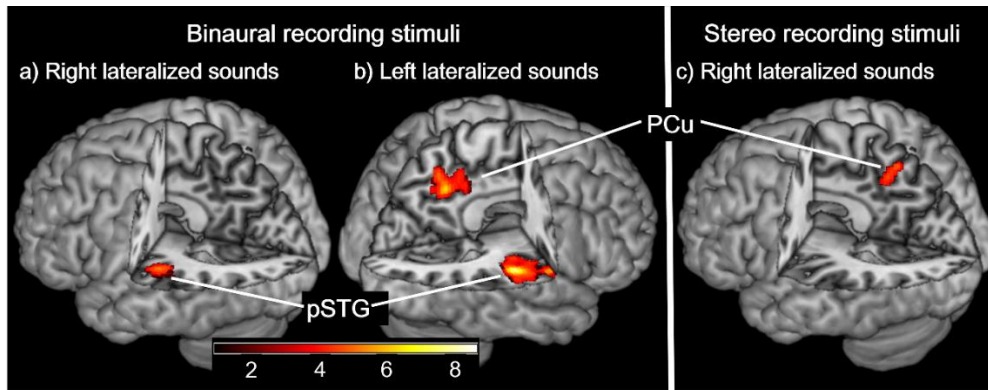


Figure 3.3. Brain regions activated to a greater extent for more lateral sound locations¹. Significant activation was rendered on a high-resolution anatomical MRI brain template and a color bar indicates t-value. (a) The left pSTG was activated to a greater extent for more right lateral ($30^\circ > 20^\circ > 10^\circ > 0^\circ$) BR stimuli. (b) The right pSTG and the right precuneus (PCu) were activated to a greater extent for more left lateral ($-30^\circ > -20^\circ > -10^\circ > 0^\circ$) BR stimuli. (c) The left PCu was activated to a greater extent for more right lateral ($30^\circ > 20^\circ > 10^\circ > 0^\circ$) stereo-recording stimuli.

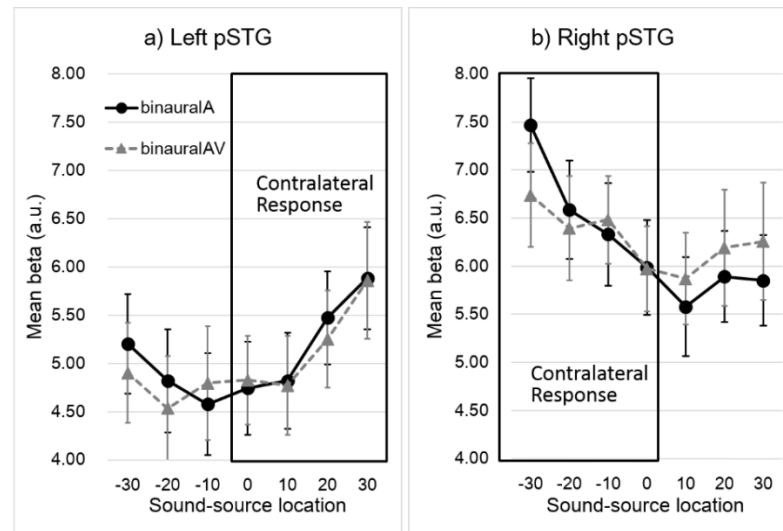


Figure 3.4. Activation of pSTG correlated with sound azimuth². Means and SEMs of beta (parameter estimates) are plotted. Both left and right pSTG responded significantly greater for contralateral BR stimuli. (a) The left pSTG responded more for right lateral sounds. (b) The right pSTG responded more for left lateral sounds and the slope was significantly steeper for the auditory-only condition than for the audiovisual condition. Binaural A, BR auditory-only; binaural AV, BR audiovisual.

¹ Callan, Callan, and Ando 2015:5, figure 3.

² Callan, Callan, and Ando 2015:6, figure 4.

Linear regression between the BOLD responses and the behavioral data

We performed linear regression analyses between the BOLD responses during the audiovisual trials and the behavioral data. With the BR stimuli, we found significant positive correlation (the height threshold $p < 0.001$ and the extent threshold $p < 0.05$ FWE corrected) in the bilateral middle temporal gyri ($x\ y\ z = 56, -12, -22$ and $x\ y\ z = -58, -14, -14$), the left middle occipital gyrus ($x\ y\ z = -38, -70, -2$), the left middle frontal gyrus ($x\ y\ z = -34, 18, 30$), the right pSTG including Heschl's gyrus ($x\ y\ z = 38, -26, 12$), the right parahippocampal gyrus ($x\ y\ z = 34, -30, -18$) and the bilateral hippocampus ($x\ y\ z = 32, -22, -20$ and $x\ y\ z = -32, -8, -26$). The scatter plots show the co-variation between brain activity (contrast estimates from the peak voxel in the right middle temporal and in the left middle occipital clusters) and the behavioral data (Fig. 3.6). As can be seen in Fig. 3.6, there is one outlier in each region. We calculated correlation without the outliers and correlations were still significant ($R = 0.76$, $p < 0.001$ in the right middle temporal gyrus, and $R = 0.65$, $p < 0.01$ in the left middle occipital gyrus). During the BR audiovisual trials, participants who were more influenced by the visual stimuli activated those regions more. With the SR stimuli, we could not find any significant correlation.

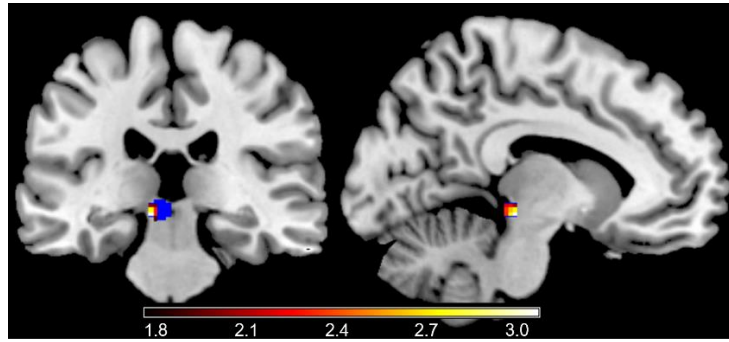


Figure 3.5. The left superior colliculus activated to a greater extent for more right lateral sound locations¹. Activation ($P < 0.05$ for display purpose) is plotted on a high-resolution anatomical MRI brain template and a color bar indicates t-value. The blue region corresponds to the left region sphere for the SVC analysis.

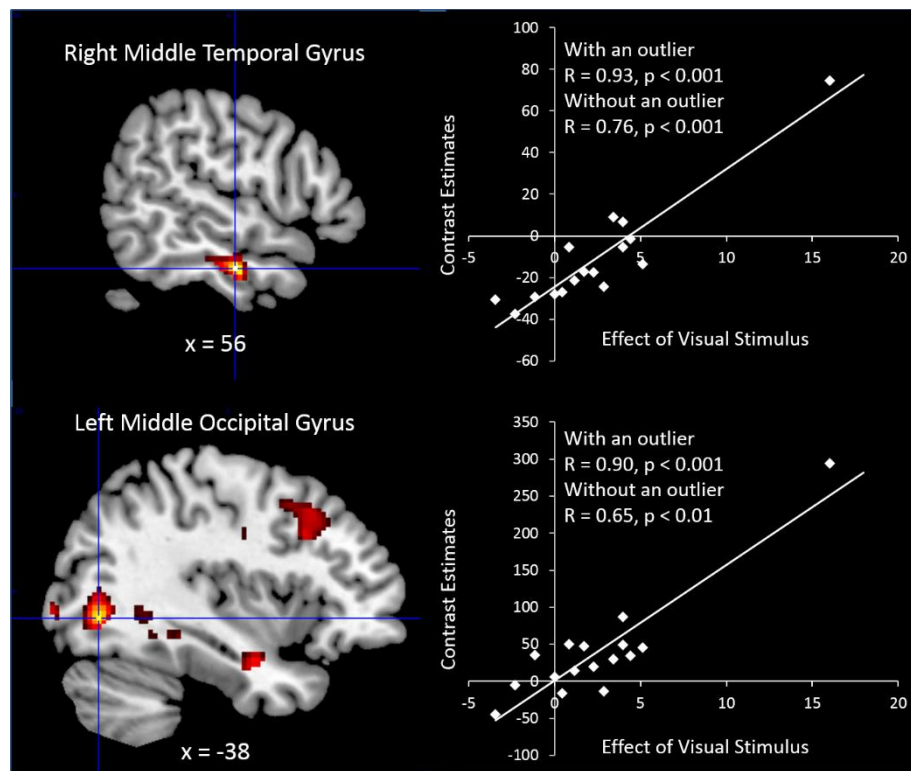


Figure 3.6. The result of the linear regression analysis between the neural responses and the behavioral data². Left: the activation maps are superimposed on a high resolution anatomical MRI brain template. Crosshairs corresponds to peak voxels from which we extracted contrast estimates in the scatter plots. Right: scatter plots illustrating the relationship between the neural responses during the audiovisual trials and the behavioral index of the ventriloquism effect. The large behavioral index indicates strong influence of visual stimuli on auditory localization. Each dot represents a separate participant.

¹ Callan, Callan, and Ando 2015:6, figure 5.

² Callan, Callan, and Ando 2015:7, figure 6.

3.4 Discussion

This is the first study to show sound azimuth-related BOLD signal changes in the human auditory cortex. The observed signal change pattern was very similar to the averaged responses of 119 neurons in the monkey auditory cortex (Werner-Reiss and Groh 2008). Neural responses did not vary much for ipsilateral sound locations and increased monotonically for more contralateral sound locations. In the audiovisual condition, centrally presented visual stimuli attenuated the monotonic change related to the sound azimuth. This attenuation of the auditory monotonic response function may be the underlying neural mechanism responsible for the ventriloquism effect.

3.4.1 Behavioral results

The significant difference between the auditory-only and audiovisual conditions confirmed that the ventriloquism effect was induced in this experimental setting. Participants' performance for left and right identification deteriorated when they were presented with visual stimuli. The behavioral effect was stronger for the SR stimuli than the BR stimuli. We suspect that the different acuity levels of auditory stimuli caused this difference. As can be seen in the Fig. 3.2, a sigmoid function of the BR stimuli was sharper than one of the SR stimuli indicating that participants identify whether the sound source was located either left or right more clearly for the BR stimuli compared to the SR stimuli. The mean percentage of “right” responses for the most left (-30°) sound was 2.5% for the BR stimuli and 4.25% for the SR stimuli. The mean percentage for the most right (30°) sound was 98.5% for the BR stimuli and 88.9% for the SR stimuli. Considering that the sensory modality with lower acuity is captured by the sensory modality

with higher acuity (Alais & Burr, 2004), it is reasonable to assume that the SR stimuli with lower acuity than the BR stimuli were more easily captured by conflicting visual information.

With the left or right forced choice task, we expected that the 0° auditory stimuli would yield near chance level (50%) right responses. The BR stimulus result was consistent with the expectation but the SR stimulus result was not. Mean percentage for the SR 0° stimuli was 26.5% indicating they were on average localized in the left side but not in the midline. In order to figure out the reason of this left lateralized responses, we inspected interaural time difference (ITD) and interaural level difference (ILD) cues of the BR and SR stimuli. We computed mean ITD values by using interaural cross-correlation and mean ILD values by ratio of the left channel RMS to the right channel RMS. For the ILD computation, the first 10 ms of the speech signals were used to prevent the reverberant sound field from affecting the estimate. The numerical values of the ITDs for the azimuth angles -30°, -20°, -10°, 0°, 10°, 20°, 30° were -0.27, -0.20, -0.12, 0.02, 0.08, 0.21, 0.35 ms, respectively for the BR stimuli and -0.19, -0.14, -0.07, 0, 0.07, 0.14, 0.19 ms, respectively for the SR stimuli. The minus ITD indicates that the left ear sound precedes the right ear sound. On the other hand, for the azimuth angles -30°, -20°, -10°, 0°, 10°, 20°, 30°, the numerical values of the ILDs were 4.80, 4.08, 2.67, 0.09, -1.77, -3.33, -4.03 dB, respectively for the BR stimuli and 1.50, 1.29, 0.77, 0.49, -0.73, -1.04, -1.25 dB for the SR stimuli. The positive ILD indicates that the left ear sound is louder than the right ear sound. The mean 0° ITDs for both the BR and SR stimuli were close to 0 ms indicating a midline position. The mean 0° ILD for the BR was 0.09 (close to 0) dB but the mean ILD for the SR was 0.49 dB indicating a left position. The positive ILD value of the SR 0° stimuli is considered a reason why they were localized to the left side on average.

3.4.2 Sound azimuth-related BOLD signal changes with the auditory-only condition

Using the externalized auditory (i.e. BR) stimuli, we found linearly increasing activity for more contralateral sounds in the pSTG. The similar neural response changes in the pSTG were not found for the SR stimuli. A parsimonious explanation of this result is that the ITD and ILD cues of the SR stimuli were less salient than ones of the BR stimuli because of the lack of the head-shadow effect. In fact, as we can tell from the mean ITD and ILD values of our stimuli, they were less salient for the SR stimuli than the BR stimuli. Nevertheless, we do not think that this is the only reason why linearly increasing activity for more left lateral sounds in the right pSTG were not found for the SR stimuli. Using N1m responses recorded by magnetoencephalography (MEG), Palomaki et al.(2005) compared individualized binaural recording (BR), non-individualized (generic) head-related transfer function (HRTF), ITD, ILD, and combined ITD and ILD (ITD + ILD) stimuli. In their study, the ITD and ILD stimuli were constructed from averaged ITD and ILD values of BR stimuli so that the ITD and ILD values of their ITD + ILD stimuli were as salient as the values of their BR stimuli. The N1m amplitude response dynamics were analyzed by subtracting the contralateral N1m amplitudes from their ipsilateral counter parts. The results of this analysis revealed significant effects of stimulus type in both the left and the right hemisphere but the dynamic range in the right hemisphere was larger than the dynamic range in the left hemisphere (Palomaki et al., 2005)). Moreover, the response dynamics in the right hemisphere reflected the amount of spatial cues in the stimuli (i.e. BR > HRTF > ITD + ILD > ITD > ILD) (Palomaki et al., 2005). Their study also reported that the correlation between the N1m amplitude and behavioral performance was much higher for the BR stimuli than for the

generic HRTF stimuli (Palomaki et al., 2005). These results indicate that more realistic auditory spatial stimuli yield a larger dynamic range of brain activity in the right auditory cortex and that the larger dynamic range is associated with better localization accuracy.

The differential neural activity in the pSTG between the BR and SR stimuli supports the hypothesis that realistic complex auditory stimuli elicit greater neural responses than artificial simple auditory stimuli in the auditory cortex (A. Callan et al., 2013; Getzmann & Lewald, 2010a; Palomaki et al., 2005). The involvement of the pSTG in processing complex sounds have been suggested by neurophysiological studies (Rauschecker, 1998; Rauschecker, Tian, & Hauser, 1995). It may also explain why previous fMRI studies using ITD and ILD cues failed to observe contralateral preferences in the auditory cortex (Brunetti et al., 2005; Woldorff et al., 1999; Zimmer et al., 2006; Zimmer & Macaluso, 2005). The right pSTG was activated to a greater extent by the more left lateralized sounds and the left pSTG was activated to a greater extent by the more right lateralized sounds. The sound azimuth-related BOLD signal change in the pSTG was revealed by the ROI analysis (Fig. 3.4). Our results provide strong support for location coding by opponent neural populations (i.e. population rate coding) in the human auditory cortex (Stecker et al., 2005). Since this experiment only investigated sound azimuth from -30° to 30° , we cannot state how BOLD signal changes for more lateralized sounds. However, based on neurophysiological work with monkeys (Werner-Reiss & Groh, 2008), we assume that neural responses are greatest for sounds at the lateral peaks but show the least modulation for sounds around the lateral peaks. Neural responses in the monkey auditory cortex showed similar activation levels between 60° and 90° (Werner-Reiss & Groh, 2008). In order to understand the whole response function in the right pSTG, an fMRI experiment to cover the whole 360° horizontal azimuth needs to

be conducted.

In addition to the pSTG, sound azimuth-related BOLD signal changes were observed in the right precuneus with the BR stimuli and in the left precuneus with the SR stimuli. The right precuneus was activated to a greater extent for more left lateral sounds and the left precuneus was activated to a greater extent for more right lateral sounds. Krumbholz et al. (2009) reported that the precuneus was involved in spatial attention shifts not only in the visual modality but also in the auditory modality and responded more strongly to attention shifts towards the contralateral than to the ipsilateral hemisphere. We suspect that the precuneus activation we found in this study may reflect spatial attention shifts caused by presenting auditory stimuli from various locations.

3.4.3 Effects of visual stimuli on auditory spatial perception

The main purpose of this study is to determine how conflicting visual stimuli affect auditory spatial perception. The auditory-only minus audiovisual contrast was used for this purpose. A significant difference was only observed in the right pSTG region with the BR stimuli (Fig. 3.4b). The results indicate that neural responses to the left lateral sounds were attenuated by observing visual stimuli presented at the frontal midline. The attenuation of the monotonic azimuthal sensitivity by spatially disparate visual stimuli was considered to reflect the neural mechanism of the ventriloquism effect such that visual spatial information perceptually captures auditory spatial information. The attenuation does not mean that monotonically increasing neural activity in the pSTG disappeared completely. If participants perceived the left lateralized sounds were located at the center every time visual stimulus was presented, one would expect to see the complete diminishment of

the monotonic neural response change. However, as can be seen in Fig. 3.2, the mean percentage of the left lateral sounds on the audiovisual condition was more than the mean percentage of the left lateral sounds on the auditory-only condition but still less than 50%. It means that the overall responses for the audiovisual stimuli were less left lateralized than the auditory-only stimuli but still left lateralized. The behavioral results agree with the neuroimaging results showing that the slope for the audiovisual sounds was significantly smaller (i.e. less left lateralized) than the slope for the auditory-only condition but still significantly positive (i.e. left lateralized).

Neuroimaging and lesion studies show that the auditory cortices in both hemispheres are involved in auditory spatial perception, however, the right hemisphere plays the dominant role in sound localization in humans (Palomaki et al., 2005). Although we found sound azimuth-related BOLD change in both left and right pSTG, the right pSTG cluster size (859 voxels) was much bigger than the left pSTG cluster size (153 voxels) (Fig. 3.3). Moreover, the effect of visual stimuli on auditory spatial processing was only observed in the right pSTG. These results also indicate the right auditory cortex dominance in sound localization. Cytoarchitectonic studies in human brain have shown greater volume of white matter in left than right Heschl's gyrus and the posterior temporal lobe indicating faster transmission and greater temporal resolution available in the left auditory cortex (Zatorre, Belin, & Penhune, 2002). A functional imaging study comparing spectral and temporal processing in auditory cortex found that temporal resolution is better in the left auditory cortical areas and spectral resolution is better in the right auditory cortical areas (Zatorre & Belin, 2001). It is well known that the ITDs and ILDs are two main mechanisms to localize a sound sources in the horizontal plane (Blauert, 1997). Spectral cues are considered to be supplemental for horizontal

localization judgment. However, greater involvement of the right pSTG in spatial processing indicate that spectral information processing is also important to determine the spatial location of the sound source in the horizontal dimension.

One limitation of this present study is that we could not compare between the neural responses to stimuli that are actually physically displaced in space versus stimuli that are displaced only as a result of the ventriloquism illusion. In this study, we could not specify trials in which they experienced the ventriloquism illusion based on their responses because we employed the left-right forced-choice task instead of the localization task. Therefore, even if visual stimulus presentation affected their perception and they perceived a 3° sound as a 20° sound, their response remains the same ('right'). Because of this reason, we could not distinguish the illusion trials from the no-illusion trials. However, our results are in line with Bonath et al.'s (2007) study that reported results of such a comparison. In their experiment, participants were asked to indicate the location of the perceived tone (central, left or right). In the case when a central sound was presented with a left or right visual stimulus, a trial where the central sound was perceived as a left or right sound was treated as an illusion trial and a trial where the sound was perceived as a central sound was treated as a no-illusion trial. By subtracting illusion trials from no-illusion trials, they found reduced activity in the planum temporale (PT) of the hemisphere ipsilateral to the visual stimulus (i.e. shifted auditory percept). They also compared left or right sounds with central sounds without visual stimuli and found enhanced activity in the PT of the hemisphere contralateral to the auditory stimuli. Their results indicate that the lateralized auditory percept is represented by asymmetrical neural responses in the PT and the laterally shifted auditory percept, in their study, resulted from a relatively enlarged responses in the

contralateral PT by reducing activity in the ipsilateral PT. The loci (i.e. the pSTG including the PT) found in this study are consistent with the loci found in their study. Moreover, the neural activity pattern change found in this study is analogous to their study. In this study, when visual stimuli were presented at the center (instead of audio in the center as in Bonath et al., 2007), the neural responses to lateral sounds that are normally asymmetric became more symmetric (in accordance to the central visual stimuli) by reducing activity in the contralateral pSTG.

In the behavioral results, the ventriloquism effect was bigger for the SR stimuli than the BR stimuli. Nevertheless, we failed to find neural response changes caused by visual stimulus presentation for the SR stimuli. The lack of significant findings could be simply due to less salient ITD and ILD cues of the SR stimuli. As we discussed earlier, we think this is unlikely because of the more dynamic response change in the pSTG for the binaural recording stimuli than the matched ITD + ILD (Palomaki et al., 2005). The pSTG, or planum temporale more specifically, has been considered as a computational hub that segregates the components of the acoustic world and matches these components with learned spectrotemporal representations (Griffiths & Warren, 2002). In our previous fMRI study, we found more enhanced activation by the BR stimuli compared to the SR stimuli in the pSTG (Callan et al. 2013). The MEG study showed artificial ITD and ILD stimuli can activate the auditory cortex but realistic auditory stimulus with spectrotemporal cues can yield a larger dynamic range of brain activity (Palomaki et al., 2005). In line with these findings, we found linearly increasing activity for more lateral sounds in the pSTG with the BR but not with the SR stimuli. We assume that direction-specific modulation in the pSTG produced by the SR stimuli was not large enough to show the significant effects of visual stimuli.

Krumbholz et al. (2005) proposed a hierarchical organization of auditory spatial processing in which analysis of simple interaural cues begins at the brainstem and more complex signal sensitivity (in their case, moving sound) emerges in the pSTG. Based on the proposal, we assumed that the ventriloquism effect for the simple SR stimuli was processed below the level of the primary auditory cortex, in the brainstem. Because the superior colliculus (SC) in the brainstem is a site for multisensory integration (Stein & Stanford, 2008), we performed additional SVC analyses at the SC. In contrast to finding in the right pSTG, linearly increasing activity in the left SC for the right lateral sounds was only found for the SR stimuli but not for the BR stimuli. Moreover, a significant difference between the slope of the auditory-only condition and the slope of the audiovisual was also found only for the SR stimuli. The result supported our assumption that the ventriloquism effect for the SR stimuli was processed below the level of the primary auditory cortex, in the SC.

3.4.4 Correlation between the neural responses and the behavioral data

We found that azimuthal representation in the right pSTG is attenuated by a spatially conflicting visual stimulus. However, we still do not know how visual processing affects auditory processing. Results of regression analyses between the BOLD responses and the behavioral data provide insights into the underlying neural mechanism. In these analyses, we found significant positive correlation in the bilateral middle temporal gyri, the left middle occipital gyrus, the left middle frontal gyrus, the right pSTG including Heschl's gyrus, the parahippocampal gyrus and the hippocampus. Most of those areas have been reported to be involved with spatial processing. The middle

occipital gyrus showed a preference for visual spatial over nonspatial processing (Renier et al., 2010), the middle frontal gyrus showed engagement in the storage of spatial information (Leung, Gore, & Goldman-Rakic, 2002), the hippocampus is involved in spatial memory (Burgess, Maguire, & O'Keefe, 2002) and the parahippocampal gyrus plays a critical role in spatial navigation (Epstein, 2008). Stronger facilitation of those areas by participants who were more influenced by spatially conflicting visual stimuli indicate that they processed visual spatial information more extensively than participants who were less influenced by spatially conflicting visual stimuli

3.4.5 Plasticity in human sound localization

The neural response changes in the pSTG observed in this study indicate the plastic nature of spatial processing in the pSTG. The plastic nature of auditory spatial perception has been demonstrated by the ventriloquism aftereffect. The ventriloquism aftereffect is a phenomenon characterized by an enduring shift in the perception of acoustic space in the absence of the visual stimulus after prolonged (20 - 30 min) exposure to a consistent audio-visual spatial disparity (Recanzone, 1998). More recent studies reported that even milliseconds of single exposure to an auditory-visual discrepancy could cause recalibration of perceived auditory space (Wozny & Shams, 2011) and that a few minutes of exposure produced consolidated and long-lasting aftereffects (Frissen, Vroomen, & de Gelder, 2012).

Since audiovisual spatial disparity was randomly changed in this study, we cannot investigate whether long-lasting neural response changes associated with the ventriloquism aftereffect occur or not. However, visually driven neural changes in the pSTG provides strong evidence that the cortical plasticity is mediated by the pSTG. A

behavioral study investigating whether ventriloquism aftereffects generalize across sound frequencies found that the aftereffects could be generalized across a four-octave range of test frequencies (Frissen, Vroomen, de Gelder, & Bertelson, 2005). In their study, the ITD dominant low frequency (400 Hz) adapter produced the aftereffect to the ILD dominant high frequency (6400 Hz) test tone and the results supported a hypothesis that the locus of recalibration was at least beyond the level of the peripheral ITD and ILD mechanisms (Frissen et al., 2005). The ITD and ILD processing pathways converge in the inferior colliculus (Konishi, 2003). Therefore, our findings showing recalibration in the pSTG for the BR stimuli and in the SC for the SR stimuli, are in line with the hypothesis provided by the behavioral study of the ventriloquism aftereffect.

3.4.6 Ventriloquism effect in relation to the auditory dorsal pathway

Like the visual system, it has been proposed that the auditory system have two functionally distinct and anatomically segregated processing streams at the cortical level (two-streams hypothesis) (Alain, Arnott, Hevenor, Graham, & Grady, 2001; Rauschecker & Tian, 2000). The two streams are dorsally or ventrally located and are considered to process location ('where') or identity ('what') information of objects, respectively. In the auditory system, the dorsal pathway emanates from the pSTG and the ventral pathway emanates from the anterior STG. Our results that show the involvement of the pSTG in auditory localization are consistent with the dorsal 'where' auditory system. However, it is unclear how the system is associated with attenuation of neural activity in the pSTG by spatially conflicting visual stimuli?

In recent years, the involvement of the dorsal pathway in

perception-action processing has been discussed. Arnott and Alain (Arnott & Alain, 2011) suggest that auditory spatial processing in the dorsal pathway may be understood as a form of action processing in which the visual system may be guided to a particular location of interest. In their account, auditory spatial processing serves to inform and orient eyes towards location of interest. They also discuss that the dorsal stream functions as a kind of filter that directs attention to particular regions in space so that the ventral stream processes particular objects. In this study, participants looked at the center of the screen in both the auditory-only and audiovisual conditions. This is to ensure that differential neural activity between the auditory-only and audiovisual conditions was not caused by simple eye positions. Attenuation of neural activity in the pSTG may indicate that visual and auditory dorsal pathways interact with each other and that the location of interest provided by vision can modify auditory localization processing.

3.5 Conclusion

Using realistic (externalized) auditory spatial stimuli, our fMRI study was able to show that bilateral pSTG were more sensitive to sound sources in contra- than in ipsilateral hemifields. ROI analyses in the pSTG revealed monotonically increasing neural responses for more contralateral locations. The results support population rate coding but not topographical place coding in the human auditory cortex. The comparison between the auditory-only conditions and the audiovisual conditions successfully demonstrated the neural basis of the ventriloquism effect as arising from attenuation of the monotonically increasing functions of sound azimuth processing in pSTG by the capture of spatially discordant visual stimuli. When locations of

auditory and visual speech signals are different but those signals were considered to be generated from the same speaker, audiovisual spatial information is integrated by attenuating auditory neural responses.

Learning in communication: Learning-related changes in neural responses to linguistic stimuli

In addition to spatial processing that we discussed in the previous section, learning is also very important for human-to-human communication. In order to understand what a speaker is saying, an individual needs to learn specific aspects of the language. In this section, we discuss how learning modifies neural responses to linguistic stimuli. We investigated differences in neural responses to native and non-native linguistic stimuli in Chapter 4 and neural response changes related to learning of nonnative linguistic stimuli in Chapter 5.

Chapter 4

Difference in neural responses to native and non-native linguistic stimuli¹

4.1 Introduction

In this chapter, we examined whether different learning levels of a language cause different neural responses in processing of the language. For this purpose, we compared neural correlates of a native phonemic contrast with neural correlates of a difficult non-native phonemic contrast.

Studies have shown that adults have difficulty discriminating non-native phonetic contrasts that are assimilated to a single phoneme in their native language (Miyawaki et al., 1975). In Japanese, English phonemes /r/ and /l/ are assimilated into one phoneme and adult Japanese speakers have difficulty identifying these phonemes even after years of exposure (Miyawaki et al., 1975). In our previous research, we compared brain activity during an English /r-l/ identification task between native English and native Japanese speakers (D. E. Callan, Jones, Callan, & Akahane-Yamada, 2004). The results show enhanced activity for Japanese over English speakers not only in brain regions thought to be involved with speech perception (anterior superior temporal gyrus, planum temporale, supramarginal gyrus) but

¹ This chapter is published as Callan, Callan, Tajima and Akahane-Yamada 2006.

also in those involved with speech production (Broca's area, anterior insula, and cerebellum). Enhanced activities in these areas were observed in perceptual training of /r-l/ phonetic contrasts by native Japanese speakers (D. E. Callan et al., 2003). In addition, enhanced activity post relative to pre training for the Hindi Dental Retroflex contrast was found to occur in inferior frontal gyrus and supramarginal gyrus ((Golestani & Zatorre, 2004); Note: supramarginal gyrus activity is reported as a subthreshold trend). The results are consistent with a hypothesis that implicates the use of auditory-articulatory mapping to facilitate perceptual identification of difficult non-native phonetic contrasts. The hypothesis is reasonable considering substantial research reports about a close relation between the neural processes of speech perception and production. Activations in motor areas involved in speech production by listening to speech stimuli have been reported by several neural imaging studies (D. E. Callan et al., 2004; D. E. Callan et al., 2003; Celsis et al., 1999; K. Watkins & Paus, 2004; Wilson, Saygin, Sereno, & Iacoboni, 2004). Studies using transcranial magnetic stimulation showed that speech listening enhanced the size of motor-evoked potentials in the tongue muscles (Fadiga, Craighero, Buccino, & Rizzolatti, 2002) and in the lip muscles (K. E. Watkins, Strafella, & Paus, 2003). Improvement of speech production accompanied by perceptual identification training of difficult non-native phonetic contrasts also indicates the close relation between speech perception and production (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997).

Previous observations (D. E. Callan et al., 2004; D. E. Callan et al., 2003; Golestani & Zatorre, 2004) that imply involvement of auditory-articulatory mapping were from studies based on non-native phonetic contrasts that differ predominantly in place of articulation (differing formant spectral transitions). It cannot be concluded that the

same auditory-articulatory mapping is used for different types of phonetic contrasts. Japanese language consists of phonetic contrasts that are distinguished solely by their durations. Many second language Japanese learners have considerable difficulty perceiving these durational contrasts (Toda, 2003). Previous neural imaging studies of durational contrasts have focused on revealing the effects of language experience on speech discrimination. These studies have found greater activity for native listeners over non-native listeners in left auditory regions (Gandour, Wong, Lowe, Dzemidzic, Satthamnuwong, Tong, & Lurito, 2002; Minagawa-Kawai, Mori, Sato, & Koizumi, 2004), as well as in left precentral gyrus, left frontal gyrus, and left fusiform gyrus (Gandour, Wong, Lowe, Dzemidzic, Satthamnuwong, Tong, & Li, 2002). Neural processes for non-native over native listeners and for speech identification were not discussed.

In this study, brain activity during Japanese short/long vowel identification and consonant identification are compared between native English and native Japanese speakers. Our study is novel in that it investigates vowel length identification for non-native over native speakers. There are two objectives for this study. The first objective is to examine whether the auditory-articulatory mapping is utilized for perception of non-native phonetic contrasts that differ in temporal aspects (i.e. duration). The Japanese vowel length contrast is difficult for native English speakers, whereas, the consonant contrast is easy for both groups. Based on the hypotheses stated above, it is predicted that non-native Japanese speakers relative to native Japanese speakers will utilize brain regions involved with articulatory-auditory mapping to facilitate perception of the vowel length contrast over that of the consonant contrast. The second objective is to investigate whether supramarginal gyrus (thought to be involved in orosensory processing) is utilized differentially for consonant identification and vowel length

identification. It is predicted that the type of auditory-articulatory mapping used to facilitate perception is dependent on the relative contribution of information important for articulatory planning control. Unlike studies for non-native spectral contrasts (D. E. Callan et al., 2004; D. E. Callan et al., 2003; Golestani & Zatorre, 2004), enhanced activity for the non-native durational contrast is not expected in brain regions involved with orosensory processing. Orosensory control is more important for consonants than vowels (Guenther, Ghosh, & Tourville, 2006). It is even more so if spectral contrasts of consonants and durational contrasts of vowels are compared. Guenther and Perkell (2004) propose that the role of supramarginal gyrus in speech production is to represent the desired articulatory movement direction in orosensory coordinates. In order to examine the second objective, the consonant contrast and the vowel length contrast were compared for native Japanese speakers. This comparison was not conducted for native English speakers to avoid possible confounds caused by differences between native and non-native phonetic contrasts.

4.2 Materials and methods

4.2.1 Subjects

Thirteen native English speakers (ages 21–49 years, 12 men and one woman) and 13 native Japanese speakers (ages 23–43 years, seven men and six women) participated in this study. English speakers had lived in Japan for 3 months to 4 years and took Japanese lessons once a week for less than 1 year. Participants were all right-handed with no hearing or speech problems and gave written informed consent for experimental procedures approved by the ATR Human Subject Review Committee.

4.2.2 Stimuli and procedure

The speech stimuli consisted of nonword minimal pairs of the form ‘ereCe’ (short vowel) and ‘ere:Ce’ (long vowel) in which ‘C’ is one of the four consonants (p, t, m, and n). A professionally trained male native Japanese speaker produced the stimuli at three speaking rates (slow, normal, and fast) that were determined by himself. The speech stimuli were recorded digitally with a sampling rate of 44.1 kHz. Signal correlated noise (SCN) stimuli that have the same amplitude envelopes as the speech stimuli but do not contain the same spectral information (see (D. E. Callan et al., 2004)) were used as a baseline condition. Two task conditions exist in this experiment; vowel length identification and consonant identification. For both tasks, short vowels (e.g. erete), long vowels (e.g. ere:te) and SCN stimuli were randomly presented. For the vowel length task, participants were instructed to press one of the three buttons corresponding to short vowel, long vowel, and noise. For the consonant identification task, participants were instructed to press one of the three buttons corresponding to stop consonant (/p/ or /t/), nasal (/m/ or /n/), and noise. Tasks were conducted in separate runs. Each participant performed one vowel length task run and one consonant identification task run. Task order was counterbalanced across participants. Stimuli were presented roughly every 2.25 s and the order was decided by using a genetic algorithm (Wager & Nichols, 2003) to optimize event-related functional magnetic resonance imaging design. Participants were asked to respond quickly to minimize differences in the hemodynamic response resulting from long response times. Participants were not given online feedback of the correctness of their responses. All participants practiced outside of the scanner on a subset of the stimuli to familiarize themselves with the tasks.

Magnetic resonance imaging data acquisition and preprocessing

For structural and functional brain imaging, Shimadzu–Marconi’s Magnex Eclipse 1.5T PD250 (Marconi Medical Systems, Cleveland, Ohio, USA) was used at the ATR Brain Activity Imaging Center. Functional T2*-weighted images were acquired using a gradient echo planar imaging sequence (repetition time=2000 ms, echo time=49 ms, flip angle=90°, field of view=192×192mm², matrix size=64×64 pixels, slice thickness=6 mm, 172 scans per run). Twenty axial slices were taken. Images were preprocessed using programs within SPM2 (Wellcome Department of Cognitive Neurology, London, UK). Images were slice time corrected, realigned, spatially normalized (voxel size 2×2×5mm³) by using a template defined by the Montreal Neurological Institute, and were smoothed using a twice voxel size (6×6×12mm³) full width at half-maximum Gaussian kernel.

4.2.3 Functional magnetic resonance imaging data analysis

Preprocessed MRI data were analyzed statistically on a voxel-by-voxel basis using the SPM2 [1/128 Hz high-pass filter, serial correlations corrected by an autoregressive AR(1) model]. The task-related neural activities were modeled with a series of events convolved with a canonical hemodynamic response function. For each participant, a vowel length-minus-consonant identification contrast (taking into account the baseline SCN condition) was assessed. For native Japanese subjects, a consonant identification minus-vowel length contrast was also assessed. The subject specific contrast images of parameter estimates were used as inputs for the second-level analysis. At the second level, a two-sample t-test for the vowel length-minus-consonant

identification contrast between English and Japanese participants was conducted as well as a one-sample t-test for the consonant identification-minus-vowel length contrast for only Japanese subjects.

4.3 Results

4.3.1 Behavioral performance

Button press responses during vowel length and consonant identification tasks were analyzed by t-test. The native Japanese subjects (mean=98%; SD=1.1) had significantly better performance than the native English subjects (mean=78%; SD=5.9) for the vowel length condition ($t(24)=12.51$; $P < 0.00001$). On the other hand, there was no significant difference between Japanese subjects (mean=95%; SD=3.8) and native English subjects (mean=96%; SD=2.8) for the consonant identification condition.

4.3.2 Brain imaging

Brain regions (see Fig. 4.1, Table 4.1) involved with identification of difficult non-native durational contrasts were determined by comparison between native English and native Japanese subjects for the vowel length-minus-consonant identification contrasts ($P < 0.005$ uncorrected; $t=2.8$; $df=24$; extent threshold=10 voxels). Figure 4.2 and Table 4.2 shows differential brain activity for the consonant identification-minus-vowel length contrasts for native Japanese subjects ($P < 0.005$ uncorrected; $t=3.05$; $df=12$; extent threshold=10 voxels). Small volume corrections for multiple comparisons were conducted in areas assumed to involve in auditory–articulatory mapping; Broca’s area, superior temporal gyrus, cerebellum (based on

(D. E. Callan et al., 2004)), planum temporale (based on (Gandour, Wong, Lowe, Dziedzic, Sathamnuwong, Tong, & Li, 2002)), and supramarginal gyrus (based on (Price, Moore, Humphreys, & Wise, 1997)).

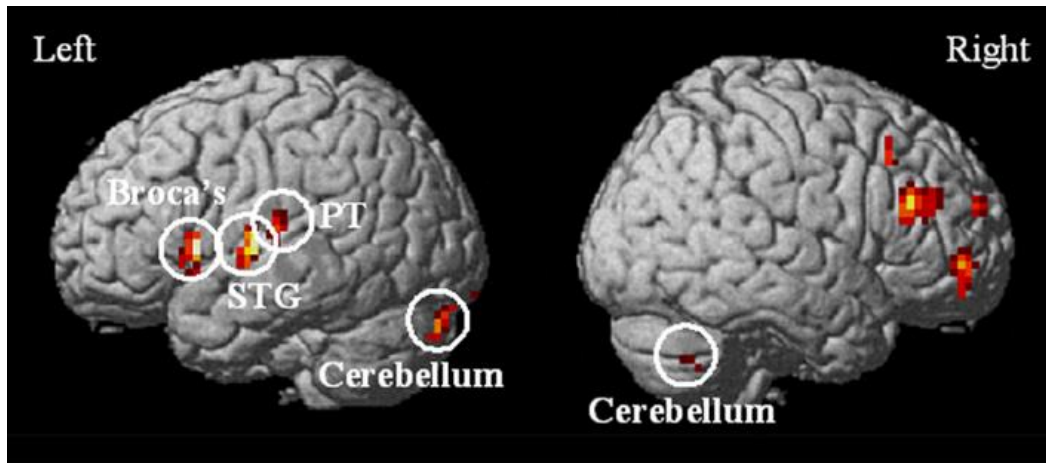


Figure 4.1. Greater differential brain activity between native English and Japanese speakers for vowel length identification relative to consonant identification ($P < 0.005$ uncorrected, $T = 2.8$, $df = 24$, extent threshold = 10 voxels)¹. STG, superior temporal gyrus. PT, planum temporale.

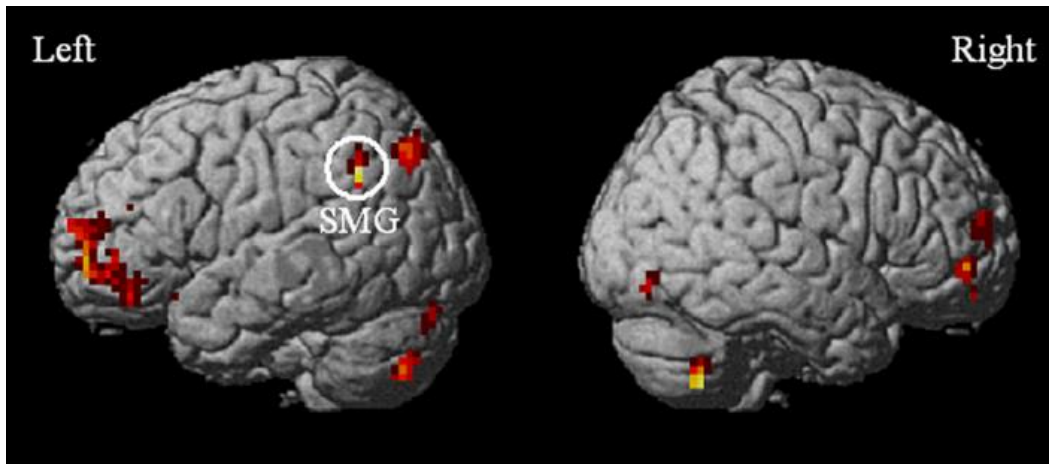


Figure 4.2. Greater differential brain activity between consonant identification and vowel length identification for native Japanese speakers ($P < 0.005$ uncorrected, $T = 3.05$, $df = 12$, extent threshold = 10 voxels)². SMG, supramarginal gyrus.

¹ Callan, Callan, Tajima and Akahane-Yamada 2006:1355, figure 1.

² Callan, Callan, Tajima and Akahane-Yamada 2006:1355, figure 2.

Table 4.1

Greater differential brain activity between native English and Japanese speakers for vowel length identification relative to consonant identification¹

Brain Region	BA	Coordinates	Peak t value	SVC center coordinates
		x, y, z (mm)		
Left superior temporal gyrus	22	-48, -9, 3	4.54 ^a	-50, 0, 0
Right inferior frontal gyrus	46	45, 30, 21	4.21	
Basal ganglia		-27, 18, 6	4.04	
Broca's area	44	-54, 12, 3	3.94 ^a	-56, 12, 8
Right middle frontal gyrus	9	6, 42, 26	3.94	
	10	30, -49, -8	3.86	
Left planum temporale	42	-48, -20, 15	3.93 ^a	-55, -11, 9
Right cerebellum		33, -60, -35	3.87 ^a	36, -57, -27
Left cerebellum		-12, -78, -26	3.34 ^a	-14, -75, -47
Right superior frontal gyrus	8	45, 22, 38	3.10	

BA, Brodmann area. ^a Small-volume corrected $p < 0.05$.

Table 4.2

Greater differential brain activity between phoneme identification and vowel length identification for native Japanese speakers²

Brain Region	BA	Coordinates	Peak t value	SVC center coordinates
		x, y, z (mm)		
Left superior frontal gyrus	8	-36, -68, 45	6.41	
	10	-27, 55, 0	5.31	
Right cerebellum		33, -60, -48	5.65	
Right middle occipital gyrus	19	33, -76, -4	5.27	
Left supramarginal gyrus	40	-45, -51, 36	5.23 ^a	-40, -46, 32
Left inferior frontal gyrus	47	-21, 35, -7	4.60	
Left cerebellum		-42, -74, -32	4.32	
Right middle temporal gyrus	21	39, -32, -1	4.12	
Right cingulate gyrus	31	27, -37, 27	4.01	

BA, Brodmann area. ^a Small-volume corrected $p < 0.05$.

¹ Callan, Callan, Tajima and Akahane-Yamada 2006: 1355, table 1.

² Callan, Callan, Tajima and Akahane-Yamada 2006: 1356, table 2

4.4 Discussion

Neural processes utilized for identification of difficult nonnative durational contrasts were studied by comparing between native English and native Japanese speakers for short/long vowel identification (difficult for English speakers) relative to consonant identification (easy for both groups). Differential activity (Fig. 4.1, Table 4.1) is present in the Broca's area, anterior superior temporal gyrus, planum temporale, the right lateral cerebellum, and the left lateral cerebellum. Brain activity in Broca's area, planum temporale, and the right lateral cerebellum are in regions very similar to those reported for positive correlation with /r-l/ perceptual identification performance in native Japanese speakers (D. E. Callan et al., 2004) as well as post relative to pretraining of /r-l/ perceptual identification (D. E. Callan et al., 2003). Activity in the anterior superior temporal gyrus is more anterior to that found for nonnative versus native /r-l/ perceptual identification (D. E. Callan et al., 2004) but is similar to that found for duration judgments of speech relative to nonspeech (Gandour, Wong, Lowe, Dzemidzic, Satthamnuwong, Tong, & Li, 2002). Activity is found in the left cerebellum, which is a region thought to be involved with encoding of durational parameters relevant to speech (Mathiak, Hertrich, Grodd, & Ackermann, 2002) and may reflect processing of the vowel length task in our study.

Consistent with findings of other studies investigating non-native phonetic perception (D. E. Callan et al., 2004; D. E. Callan et al., 2003), the results of this study show activity in regions involved with speech perception and speech production (Fig. 4.1, Table 4.1). Broca's area is known to be involved in speech production and has also been implicated in certain speech perception tasks (K. Watkins & Paus, 2004). The planum temporale is conjectured to be involved with mediating

auditory–motor transformation (J. E. Warren, Wise, & Warren, 2005). Activity in the right lateral cerebellum is in a region shown to somatotopically represent the lips and tongue (Grodd, Hulsmann, Lotze, Wildgruber, & Erb, 2001).

Unlike studies for perception of non-native spectral contrasts (D. E. Callan et al., 2004; D. E. Callan et al., 2003; Golestani & Zatorre, 2004), differential activity in the supramarginal gyrus was not found for identification of difficult non-native durational contrasts. On the other hand, enhanced activity in the left supramarginal gyrus was found for consonant over vowel length identification by native Japanese speakers (Fig. 4.2, Table 4.2). Behavioral data indicate that both tasks were easy for Japanese speakers (95% consonant identification; 98% vowel length identification). Left supramarginal gyrus activation was reported in a phonological short-term memory task relative to a rhyming task (Paulesu, Frith, & Frackowiak, 1993) and in listening to syllable sequences with one deviant stimulus relative to syllable sequences with all identical stimuli (Celsis et al., 1999). Those results imply that the left supramarginal gyrus has a role for manipulating phonological segments in working memory. The left supramarginal gyrus activation is also found in manipulating nonspeech (hum) segments (Gelfand & Bookheimer, 2003) and in judging Thai tones (Gandour, Wong, Lowe, Dzemidzic, Satthamnuwong, Tong, & Li, 2002). The common factor for all those results is detecting frequency spectral change. Human voice spectrum is formed by vocal fold vibration cycles and vocal tract shape (the source filter theory) (Fant, 1960). Difference in consonants is mainly caused by vocal tract shape and difference in pitch (F0) is mainly caused by vocal fold vibration cycles. In our experiment, the left supramarginal gyrus was less active for vowel length identification than consonant identification. Consistent with our results, the enhanced activity in the left supramarginal gyrus was not

found in judging vowel lengths in Thai by native Thai speakers, although it was found in judging Thai tones (Gandour, Wong, Lowe, Dziedzic, Satthamnuwong, Tong, & Li, 2002). These results suggest that the role of the left supramarginal gyrus is not merely manipulating phonological segments but integrating phonological segments and articulator position information.

There are several issues that have to be kept in mind when interpreting our results. Activity in Broca's area can be caused by task difficulty and corresponding reaction time differences (cognitive control) (Novick, Trueswell, & Thompson-Schill, 2005). Background scanner noise could affect easy and difficult tasks differently. Previous studies, however, show greater activity correlated in Broca's area and cerebellum with better non-native identification performance (D. E. Callan et al., 2004) as well as after successful identification training (D. E. Callan et al., 2003) suggesting that results are not just due to task difficulty. Unequal sex ratio between groups may affect our results, although we did not find any sex effects in our previous study (D. E. Callan et al., 2003).

4.5 Conclusion

The results of this experiment show differential activity for non-native over native speech perception of a duration phonetic contrast in brain regions involved with auditory-articulatory mapping. Enhanced activity in supramarginal gyrus for consonant identification over vowel length identification implies that the supramarginal gyrus plays a more important role for phonetic contrasts differing in place of articulation than in vowel duration. Although comparisons of native and non-native linguistic processing allow us to speculate on whether neural representations of learned linguistic stimuli are different from neural

representation of not-learned linguistic stimuli, it does not prove that learning causes modification of neural representations. In the next chapter, we investigate whether learning can modify neural responses to linguistic stimuli.

Chapter 5

Neural response changes related to learning of non-native linguistic stimuli¹

5.1 Introduction

In this chapter, learning-related changes in neural responses to linguistic stimuli were examined as a part of functional magnetic resonance imaging (fMRI) study for reading. We compared brain activity before and after learning non-native letters (phonograms).

Letters are simple two-dimensional visual forms that have phonological information. For literates, reading letters is automatic and it is very difficult to prevent the process. The Stroop (1935) color-word test is a good example demonstrating the automaticity of the reading process. On the other hand, non-native letters for which pronunciation is unknown are just meaningless symbols. There is no automatic visual to phonological conversion when you see non-native letters. In order to determine brain regions involved with the process of visual to phonological conversion, one can investigate how neural activity for perceiving non-native letters changes as a result of learning the corresponding pronunciations.

There are quite a few functional imaging studies investigating neural processes underlying letter perception. Letter perception can be separated into two sub-processes: visual form processing and

¹This chapter is published as Callan, Callan, and Masaki 2005.

grapheme-to-phonological conversion processing. To extract letter specific visual form processing, researchers compared letters with visual stimuli that belong to different categories. Letters were compared to faces (Allison, Puce, Spencer, & McCarthy, 1999; Puce, Allison, Asgari, Gore, & McCarthy, 1996; Tarkiainen, Cornelissen, & Salmelin, 2002), digits (Polk & Farah, 1998; Polk et al., 2002), checkerboards (Cohen et al., 2002), single geometric shapes (Tagamets, Novick, Chalmers, & Friedman, 2000), and objects (Joseph, Gathers, & Piper, 2003; Price & Friston, 1997; Sargent, Zuck, Levesque, & MacDonald, 1992). Many of these studies reported that letters activated left fusiform gyrus (BA37) more than other categories (Cohen et al., 2002; Polk & Farah, 1998; Polk et al., 2002; Puce et al., 1996; Tagamets et al., 2000; Tarkiainen et al., 2002). The activation in left BA37 is often reported by functional neuroimaging experiments comparing reading words with the resting state (Beauregard et al., 1997; Brunswick, McCrory, Price, Frith, & Frith, 1999; Fiez, Balota, Raichle, & Petersen, 1999; Paulesu et al., 2000; Wagner et al., 1998). Cohen et al. (2000; 2002) studied hemifield-independent visual word form processing and labeled the middle portion of left fusiform gyrus as the visual word form area based on their findings and the clinical studies for patients with alexia. Since activation in this area is observed not only for visual word processing but also for other processing (e.g., object recognition, naming, auditory words processing), appropriateness of this labeling is still under debate (Cohen & Dehaene, 2004; Price & Devlin, 2003, 2004). Reported loci for letter string processing and for visual word processing are close to each other but Cohen et al. (2002) reported that left BA37 is activated more for words than for consonant strings.

Flowers et al. (2004) have reported that activation in left BA37 is not just activated by presence of letters but by attending to letters

within a linguistic context. In their experiment, the same set of visual stimuli (single letters or symbols in either black or white color) was presented in separate sessions. For each session, subjects were told to respond to a specific target (either letter, symbol, or color) by pressing a button. According to their result, left BA37 is more activated when subjects responded to letters than to symbols or to colors even though subjects saw the same visual stimuli for all sessions. Furthermore, Joseph et al. (2003) suggest that processing in left BA37 is not category specific but is shared by both letters and objects. Results from the cognitive conjunction analysis also showed that left BA37 is conjointly activated when subjects named visually presented words, objects, letters, or colors (Price & Friston, 1997). Since reading is a postnatally acquired skill that has only gained widespread use since the 20th century, it is very unlikely that a certain cortical region evolved that is dedicated solely for reading alone. However, in non-human primates, the ventral visual pathway including the posterior inferior temporal-occipital region is activated by recognition of familiar objects (Ungerleider & Mishkin, 1982). Moreover, experience with visual stimuli has been found to lead to the development of neuronal assemblies in perirhinal cortex, a region in anterior medial inferior temporal cortex (Erickson, Jagadeesh, & Desimone, 2000). It is plausible to hypothesize that for humans, neuronal assemblies within BA37 (thought to be an auditory visual association area (Talairach & Tournoux, 1988)) develop to preferentially respond to letters based on experience. Phonological processing in reading has been studied in various ways. Based on lesion studies, left angular gyrus has been considered a site associating visual and auditory (phonological) information in reading (Damasio & Damasio, 1983; Dejerine, 1891, 1892; Geschwind, 1965; Henderson, 1986). Geschwind (1965) hypothesized a neural route connecting the visual cortex and

Wernicke's area through left angular gyrus in reading and this model is called "Wernicke-Geschwind model". Angular gyrus is located between occipital, temporal, and postcentral regions of the hemisphere. Visual, auditory, vestibular, cutaneous, and proprioceptive sensations overlap in this area (Luria, 1973). It is thought that angular gyrus has an important role for associating heard names and seen or felt objects (Geschwind, 1965). Therefore, it is plausible to hypothesize that angular gyrus is a site involved with associating visual and auditory (phonological) information in reading. However, many results from functional imaging studies do not agree with this theory. There are no specific cortical regions that are consistently reported by functional imaging studies, but there are several regions that are reported by multiple studies. They are left temporoparietal cortex including supramarginal gyrus and angular gyrus (Fujimaki et al., 1999; Kuriki, Hirata, Fujimaki, & Kobayashi, 1996; Law et al., 1991; Lee et al., 2003; Paulesu et al., 1993; Petersen, Fox, Posner, Mintun, & Raichle, 1988; Pugh et al., 1996), left superior temporal gyrus (Fujimaki et al., 1999; Kuriki et al., 1996; Paulesu et al., 1993; Petersen et al., 1988; Pugh et al., 1996; Sergent et al., 1992), and left inferior frontal lobe including Broca's area and insula (Fujimaki et al., 1999; Paulesu et al., 1993; Petersen et al., 1988; Pugh et al., 1996; Sergent et al., 1992). Poeppel (1996) criticized the phonological processing studies' non-overlapping results and pointed out three possible causes: (1) task decompositions were not sufficient; (2) distinct aspects of phonological processing were investigated; and (3) there are limitations in cognitive subtraction by PET. Specifically, the studies reviewed (Demonet et al., 1992; Paulesu et al., 1993; Petersen et al., 1988; Sergent et al., 1992; Zatorre, Evans, Meyer, & Gjedde, 1992) often subtracted activity from different types of tasks. Moreover, the influence of subtle variations of stimuli presentation and the uncontrollable automatic process of linguistic

stimuli were real so suggested to be the case for the inconsistency of results from various brain imaging studies (Indefrey et al., 1997; Price, 2000; Price et al., 1994; Sargent et al., 1992). In addition to these causes, we suspect that the type of orthography might be related to the inconsistency in the results of brain imaging studies investigating phonological processing involved with reading. Many phonological processing studies use whole words as stimuli. Phonological processing of words is more complicated than of letters because it is not only a bottom-up (sublexical grapheme-to-phonological) process but also a top-down (lexical knowledge-to-phonological) process. This top-down influence on phonological processing is significant in a deep orthography such as English because grapheme-to-phonological conversion is not a one-to-one process (Wydell, Vuorinen, Helenius, & Salmelin, 2003). Behavioral studies comparing effects of word length and word frequency have shown that word length has stronger influence on naming latency than word frequency in Dutch and Italian, in which grapheme-to-phoneme mappings is straightforward (shallow orthography). The results suggest that bottom-up processing has a stronger influence on reading in shallow orthographies than in deep orthographies (Bates, Burani, D'Amico, & Barca, 2001; Hudson & Bergman, 1984). Paulesu et al. (2000), using positron emission tomography (PET), compared the Italian (shallow orthography) and English (deep orthography) students during reading words and non-words. They reported a language-related difference: left superior temporal regions were activated more for Italian readers and left posterior inferior temporal region and anterior inferior frontal gyrus were activated more for English readers. This indicates that the orthography type in a particular language environment has a substantial influence on neural processes in reading. One may assume that the top-down process can be avoided if single letters are used as stimuli.

However, naming an English alphabet is not a simple grapheme-to-phonological conversion because the name is not the sound representation of the letter (e.g., [kei] is a name and /k/ is an associated phoneme of a letter 'k'). Therefore, if one wants to find out cortical areas that are responsible for a simple grapheme-to-phonological conversion, single letters from a shallow orthography have to be used as stimuli.

Japanese, Korean, and Thai use shallow orthographies. More precisely, Japanese and Korean use both shallow orthography (phonogram) and deep orthography (ideogram). In our experiment, phonograms of these three languages were used as visual stimuli. Japanese who could not read either Korean or Thai served as subjects. The differences of neural activity for native versus nonnative orthographies and for pre- versus post-training were investigated. Two types of non-native orthographies (one trained and one not trained) were used in order to control for differential activity that might result from practice effects for the working memory task as well as varying physiological and neurological noise that may occur during the post-test session versus the pre-test session. Half of the subjects learned Korean and the other half learned Thai. The use of two orthographies that subjects were trained on instead of one allows greater generalizability of the results (data were analyzed based on trained or non-trained condition irrespective to orthography type) helping to avoid confounds associated with identifying brain activity that may be specific to the processing of a certain writing system.

This experiment was designed to overcome some of the problems associated with experiments investigating phonological processing aspects of reading. First, simple grapheme-to-phonological conversion was investigated by using single phonograms from a shallow orthography as visual stimuli. This avoids potential top-down linguistic

processing of the stimuli instead of bottom-up processing evoked by the stimulus. It may be more appropriate to call this process as character-to-phonological conversion because grapheme in many languages such as English can be a cluster of letters (e.g., -ough in cough or tough). Second, variance caused by different task demands was avoided by employing the same working memory task (two-back identification task) for pre- and post-tests. This avoids differences in brain activity caused by differential processing of the task rather than to the process under investigation. Although in principle, accomplishing the two-back identification task with letters or words does not necessitate reading processes, it has been claimed that when the task involves encoding of the stimuli it is difficult to ignore stimulus-specific properties (Tagamets et al., 2000). Tagamets et al. (2000) justified the use of a one-back test to investigate neural processes underlying reading by citing studies demonstrating that brain regions involved with processing stimulus-specific properties are enhanced when active encoding is required to accomplish the task (Baker, Frith, Frackowiak, & Dolan, 1996; Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1991; Courtney, Ungerleider, Keil, & Haxby, 1996; Haxby et al., 1994). The Stroop (1935) effect attests to the difficulty in ignoring word information when the task is to name the color of the word. Third, variance caused by different visual features was eliminated because the same visual stimuli were used for both pre- and post-tests. This avoids differences in brain activity caused by subtle differences that may exist between stimuli. With the exception of physiological and neurological noise, the only difference between pre- and post-tests is expected to be caused by the acquired phonological knowledge associated with the newly learned non-native letters. Moreover, physiological and neurological noise were minimized by keeping the fMRI experiment time short. In general, learning

experiments take a long time (i.e., weeks, months, or years) because it takes a long time for subjects to learn new knowledge completely. Since pre- and post-fMRI scans are obtained at very different times, their physiological and neurological noise levels are very different. In our experiment, subjects only learned five letters, so the whole experiment (pre-test, training phase, and post-test) took less than 1 h and all data were acquired without the subject leaving the MRI scanner. Because of the small number of letters and short duration of training, neural process of newly learned phonograms might be very different from native phonograms. We do not expect that this short experience leads to long-term organization of neuronal assemblies in BA37 but do expect existent neural processes involved with character-to-phonological mapping to be highly activated for newly learned phonograms.

Lee et al. (2003) performed a similar learning experiment using Korean and Japanese words to investigate “the changes in brain activation patterns associated with the learning of foreign words”. Based on the results of their study, they attributed left angular gyrus activation as reflecting newly acquired character-to-phonological conversion processing. Unlike Lee et al.’s (2003) study, the purpose for our study is more specified to identify brain regions involved with visual form processing and character-to-phonological conversion processing in letter perception. Moreover, our study differs from theirs in many respects: (1) they used ‘words’ as stimuli instead of ‘letters’, so it was difficult for them to prevent top-down influence on phonological processing and to distinguish character-to-phonological association from written word-to-phonological or written word-to-semantic association; (2) the duration of their experiment was 16 days instead of 1 h, so direct comparisons between before and after learning was more difficult than our study, in fact, they did not report these comparisons in their paper; and (3) their task was different for

the various visual stimuli (subjects were asked to read silently visual stimuli if they could and just looked at them if they could not read them); our study employed the same two-back task for the different visual stimuli. One potential limitation in the Lee et al. (2003) study that is avoided in our study is their use of different tasks for the various visual stimuli. As was discussed above, it may be somewhat difficult to discern brain activity underlying a certain process when different tasks are employed for the various conditions used to determine the process of interest. This may be the reason why Lee et al. (2003) did not find left angular gyrus activation in “silent reading” of Japanese words compared with “looking” at non-trained Korean words although they found the activation in “silent reading” of Japanese words compared with “silent reading” of trained Korean words during the first fMRI experiment on the second day of the learning phase.

Comparisons were planned to investigate brain activity underlying two different aspects of letter processing. The first is the comparison between native (Japanese) and non-native (Thai and Hangul) letter processing. This comparison was planned to assess the neural activity associated with letter-specific visual form processing. Because the non-native orthographies are meaningless symbols before training and subjects have much more experience with native orthographies, it is hypothesized that left fusiform gyrus (BA37) should show significant differential activity for this comparison if the area is functionally tuned for visual form processing of letters. The second is the comparison of post-relative to pre-tests for trained and non-trained phonograms. This comparison was planned to assess the change in neural activity in learning new phonograms and thus identify brain regions involved with character-to-phonological conversion. A further comparison of post- minus pre- for the trained non-native phonograms relative to the native and non-trained non-native phonograms controls

for differential activity that may result from practice effects as well as varying physiological and neurological noise. It is hypothesized that the newly acquired character-to-phonological mappings cannot be ignored in carrying out the two-back task. In support of this hypothesis is the finding that stimulus-specific areas are activated when the task requires encoding of the stimulus (Tagamets et al., 2000). Therefore, it is expected that the brain regions involved with the character-to-phonological conversion processes (left angular gyrus) should show greater activity after relative to before learning.

Moreover, psychophysiological interaction (PPI: Friston et al., 1997; Gitelman, Penny, Ashburner, & Friston, 2003) analysis was performed to assess the functional integration of left BA37 with other areas in relation to visual form processing of native letters. An additional PPI analysis was performed to assess the functional integration of left BA39 with other areas in relation to character-to-phonological processing of newly learned phonograms. For character-to-phonological processing, we were especially interested in the functional integration of left BA39 with primary visual cortex as well as with superior temporal gyrus based on the Wernicke–Geschwind model (Geschwind, 1965).

5.2 Methods

5.2.1 Subjects

Eleven Japanese volunteers (ages 22–37 years, 8 males and 3 females), who had no previous knowledge of the Korean or Thai, participated in this study. All subjects were right-handed and scored above +80 on the Edinburgh Handedness Inventory (Oldfield, 1971). Subjects gave written informed consent for experimental procedures, approved by the

ATR Human Subject Review Committee.

5.2.2 Overall experimental design

This study consists of three parts; pre-test, training phase, and post-test. The procedures for the pre- and post-tests were identical

Pre- and Post- tests

Subjects performed a two-back identification task during pre-and post-tests. Three types of letters were used as stimuli: Japanese, Thai, and Korean. Japanese was a native language and Thai and Korean were non-native languages for all subjects. All letters were phonograms, but subjects did not know how to pronounce either Thai or Korean phonograms. Five letters from each language (Fig. 5.1) were used and corresponded to five vowels (/i/, /e/, /a/, /o/, and /u/). Subjects learned either Thai or Korean letters during the training phase.

Trials were blocked by condition and there were three conditions: native (Japanese), trained non-native (Thai/Korean), and non-trained non-native (Korean/Thai). Each condition appeared four times in each session. Subjects did one session as a pre-test and one session as a post-test. Each task condition was alternated with a baseline condition in which just a fixation mark was presented. Each task condition lasted 24 s and the baseline condition lasted 9 s. The order of task conditions was randomized across sessions. In each block, letters were presented randomly twelve times at a rate of 1 every 2 s (1.5 s presentation of a letter followed by 0.5 s presentation of a fixation mark). Subjects responded to each stimulus presentation by pressing one of the two buttons on a response box held in their left hand. They were instructed to press the left button when the letter was identical to the stimulus

presented two times prior and to press the right button when it was not. Subjects only used their left thumb to press buttons. They did not have to respond for the first two stimuli, so the response started after the third stimulus presentation. To reduce brain activity caused by confusion to the task, subjects practiced the task until they could perform with over 70% accuracy prior to entering the MRI scanner. During the imaging session, stimuli were projected onto a screen at the head of the scanner bed via an LCD projector and subjects viewed the screen via a mirror attached to the head coil.

	/i/	/e/	/a/	/o/	/u/
Japanese	い	え	あ	お	う
Korean	이	에	아	오	우
Thai	อี	เอ	อา	โ	อุ

Figure 5.1. Phonograms used as visual stimuli¹.

Training Phase

Subjects learned either five Korean or five Thai phonograms during the training phase. The type of letter that they learned during the training phase was counterbalanced between subjects. Stimuli were native (Japanese) and one of the non-native languages that was the same used in pre- and post-tests. Trials were blocked by condition and there were

¹ Callan, Callan, and Masaki 2005:556, figure 1.

three conditions: reading Japanese letters (RJL), learning non-native letters (LNL), and reading nonnative letters (RNL). For both the reading conditions (RJL and RNL), a letter was presented above the fixation mark that was presented in the center of the screen for 1.5 s. Subsequently, a square box was presented under the fixation mark for 1.5 s. Subjects were instructed to read out the letter as soon as they saw the square box or guess and produce the sound of the letter if they did not know how to pronounce it. Letters were presented randomly 5 times at a rate of 1 every 4 s. For the learning condition (LNL), a non-native letter was presented above the fixation mark, and 1.5 s later, a native letter that corresponded to pronunciation of the non-native letter was presented under the fixation mark for 1.5 s. For this condition, subjects were instructed to read out the native letter and try to associate the non-native letter with the speech sound. Letters were presented randomly 10 times at a rate of 1 every 4 s. Each letter was presented twice in each block. The order of the condition was always the same: RJL–LNL–RNL. This cycle was repeated four times in each session and subjects performed three sessions to learn non-native letters perfectly. Speech production responses were monitored by microphone attached to the head coil and recorded on a compact disc for later analysis.

5.2.3 MRI data acquisition and preprocessing

For structural and functional brain imaging, Shimadzu-Marconi's Magnex Eclipse 1.5 T PD250 was used at the ATR Brain Activity Imaging Center. Functional T2*-weighted images were acquired using a gradient echo planar imaging sequence (TR = 3000 ms, TE = 49 ms, flip angle = 90°, field of view = 192 × 192 mm, matrix size = 64 × 64 pixels). Thirty consecutive axial slices (thickness, 5 mm) covering the whole cortex and cerebellum were acquired. Before the acquisition of

functional images, T2-weighted anatomical images were acquired in the same plane as the functional images (voxel size = $0.75 \times 0.75 \times 5$ mm). T1-weighted anatomical images (voxel size $1 \times 1 \times 1$ mm) were also acquired.

Images were preprocessed using programs within SPM2 (Wellcome Department of Cognitive Neurology, University College, London). Differences in acquisition time between slices were accounted for, movement artifact was removed, and images were spatially normalized (voxel size $2 \times 2 \times 4$ mm) by using a template defined by the Montreal Neurological Institute (MNI) and were smoothed using a 10-mm FWHM Gaussian kernel.

5.2.4 fMRI data analysis

Preprocessed MRI data were analyzed statistically on a voxel-by-voxel basis using SPM2. The time series for each voxel was high-pass filtered to 1/198 Hz, serial correlations were corrected by an autoregressive AR(1) model, and global signal changes were removed by scaling. The task-related neural activities were modeled with a box-car function convolved with the hemodynamic response function. For each subject, a native minus non-native contrast and post- minus pre-test contrasts (taking into account the baseline condition implicitly in the design) for each language were assessed. Then the subject-specific contrast images of parameter estimates were used as inputs for the second (random effect) level analysis. At the second level, one-sample t tests were conducted and a threshold of $P < 0.05$ (random field theory correction for multiple comparisons) was employed.

For the computation of psychophysiological interaction (PPI) regressors representing the interaction of the psychological context with inputs from a specified region's activity, we used the process of

signal deconvolution embedded in SPM2. The first eigenvariables were calculated for the voxels (sphere with 10 mm radius) of interests (left BA37 and left BA39). PPI regressors were calculated using a native minus non-native contrast for the pre-test for BA37 and using the trained non-native condition minus the baseline in pre- and post-tests for BA39. Those regressors were inserted as user-specified regressors in new session-specific design matrices and the contrast for the PPI was calculated for each subject. Then the subject-specific contrast images were used as inputs for the second level analysis. At the second level, one-sample t test was conducted for BA37 and paired t test with non-sphericity correction was conducted for BA39. A threshold of $P < 0.001$ (uncorrected) with a $P < 0.05$ (uncorrected) for cluster extent was employed. Additionally, small volume corrections ($P < 0.05$ corrected, sphere with a radius of 6 mm) were performed for the priori regions of interest in primary visual cortex and superior temporal gyrus.

5.3 Results

5.3.1 Behavioral performance

Button press responses during pre- and post-tests were analyzed using a 3 (language: native, trained non-native, and non-trained nonnative) $\times 2$ (session: pre- and post-tests) within subjects analysis of variance (ANOVA). The mean percentages of correct responses and the mean reaction times (RTs) are plotted in Fig. 5.2. Because of a machine failure, one subject's data were lost. Subjects responded significantly more accurately ($F(1,9) = 13.04$; $P < 0.01$) and faster ($F(1,9) = 13.21$; $P < 0.01$) in the post-tests. There is no significant interaction for accuracy (i.e., improvements in accuracy were not significantly different for different language conditions), but there is a significant

interaction for RT ($F(2,18) = 12.33$; $P < 0.01$) (i.e., improvements in RT were significantly different for different language conditions). The simple effects of session for RT were assessed for each language. RTs for native and non-trained nonnative conditions were significantly reduced in the post-tests ($F(1,19) = 12.49$, $P < 0.01$ for native; and $F(1,19) = 53.23$, $P < 0.01$ for non-trained non-native). These analyses were also significant with the Geisser–Greenhouse correction, for which the critical $F(1,9) = 10.6$. On the other hand, there was no significant difference in RTs between pre- and post-tests for the trained non-native condition. In the post-tests, RTs were significantly shorter for the native condition relative to the trained non-native condition ($t(9) = 4.55$; $P < 0.002$), as well as for the non-trained non-native condition relative to the trained non-native condition ($t(9) = 4.31$; $P < 0.002$).

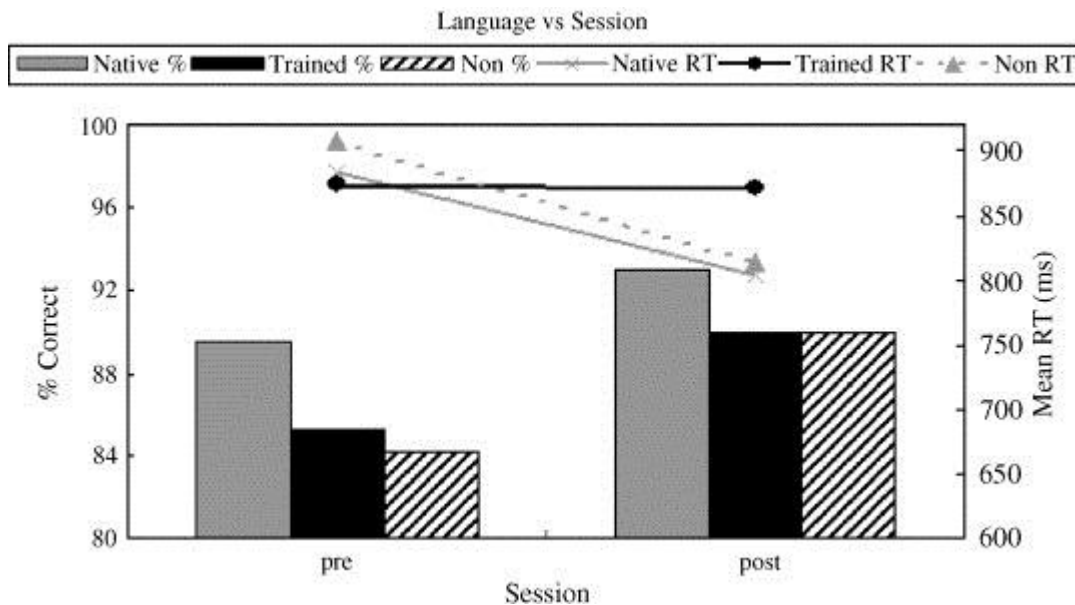


Figure 5.2. The mean percentages of correct responses and the mean reaction times (RTs)¹. Bar graphs indicate the mean percentages with the scale on the left. Line graphs indicate the RTs with the scale on the right. Native = native language (Japanese). Trained = trained nonnative language (Korean/Thai). Non = nontrained nonnative language (Thai/Korean).

5.3.2 Brain imaging

Brain regions involved with letter-specific visual form processing were determined by comparison of native with non-native orthographies. The hypothesis tested is whether left BA37 is activated more for the native orthography (Japanese) than for the non-native orthographies (either Korean or Thai). Neither contrast revealed significant differential activity in left BA37 using random field theory correction or even a region of interest analysis without correction ($P > 0.05$ uncorrected). The region of interest analysis was conducted in left fusiform gyrus BA37 (10 mm sphere at Talairach coordinates (TC: -40, -59, -9) (Talairach and Tournoux, 1988). This voxel was chosen because it

¹ Callan, Callan, and Masaki 2005:557, figure 2.

locates center of the foci of left fusiform gyrus BA37 reported by numerous investigators [e.g., Cohen et al., 2000 (-42, -57, -6); Cohen et al., 2002 (-39, -57, -9); Flowers et al., 2004 (-42, -71, -9); Joseph et al., 2003 (-41, -48, -6); Polk et al., 2002 (-37, -38, -7); Price and Friston, 1997 (-44, -62, -16); Puce et al., 1996 (-40, -66, -17); Tarkiainen et al., 2002 (-37, -70, -12)]. Additional analyses were also conducted in which the non-native orthography conditions were combined for the pre-test, post-test, and the pre- and post-tests together. There were no significant differences for these contrasts assessed using the random field theory correction for multiple comparisons or the region of interest analyses ($P > 0.05$ uncorrected). It should be noted, however, that both the native and non-native conditions showed considerable activity in fusiform gyrus BA37 when compared to the baseline condition (peak at TC (-40, -67, -13), $z = 4.89$, $P < 0.05$ corrected, for the native condition, and peak at TC (-44, -68, -10), $z = 5.65$, $P < 0.05$ corrected, for the non-native condition).

Brain regions involved with automatically induced character-to-phonological (i.e., simple grapheme-to-phonological) conversion were determined by comparing the pre- and post-tests for each language condition separately. There were no significant differences between the pre- and post-tests for both the native and non-trained non-native conditions. For the trained non-native condition, there was a significant activation ($z = 5.30$, $P < 0.05$ corrected) in left BA39 (TC: -40, -59, 18) (Fig. 5.3). It is in a region identified by several studies as angular gyrus [e.g., Bookheimer, Zeffiro, Blaxton, Gaillard, and Theodore (1995) (-44, -66, 32); Demonet et al. (1992) (-48, -52, 28); Howard et al. (1992) (-50, -48, 8); Joubert et al. (2004) (-40, -53, 32); Lee et al. (2003) (-56, -66, 22); Price et al. (1994) (-52, -38, 24); Small et al. (1996) (47, 58, 19)]. An interaction analysis comparing post- minus pre- for the trained non-native condition

relative to the native and non-trained non-native conditions showed significant differential activity for voxels identified in the above analysis within left angular gyrus ($z = 2.24$, $P < 0.05$ corrected).

Additionally, psychophysiological interaction (PPI: Friston et al., 1997; Gitelman et al., 2003) analyses were performed to assess the functional integration of left BA37 with other areas on visual form processing of native letters and one of left BA39 with other areas on character-to-phonological processing of newly learned phonograms. For the left BA37 PPI analysis, the peak voxel (TC: -44, -68, -10) in left BA37 during the two-back identification task (including both native and non-native language conditions) compared to baseline was used as the center voxel of interest. Regions more strongly connected to left BA37 in the native condition compared to the non-native conditions were right anterior cingulate (BA24) and right precentral gyrus (BA6) (Table 1). For the BA39 PPI analysis, the peak voxel (TC: -40, -59, 18) in the pre- and post-test comparison for the trained non-native condition was used. Table 2 shows the regions more strongly connected to left BA39 during processing non-native phonograms after training compared to before training. The results indicated that left BA39 was more strongly connected to primary visual cortex (BA18) and superior temporal cortex (BA22) in the trained non-native condition after training.

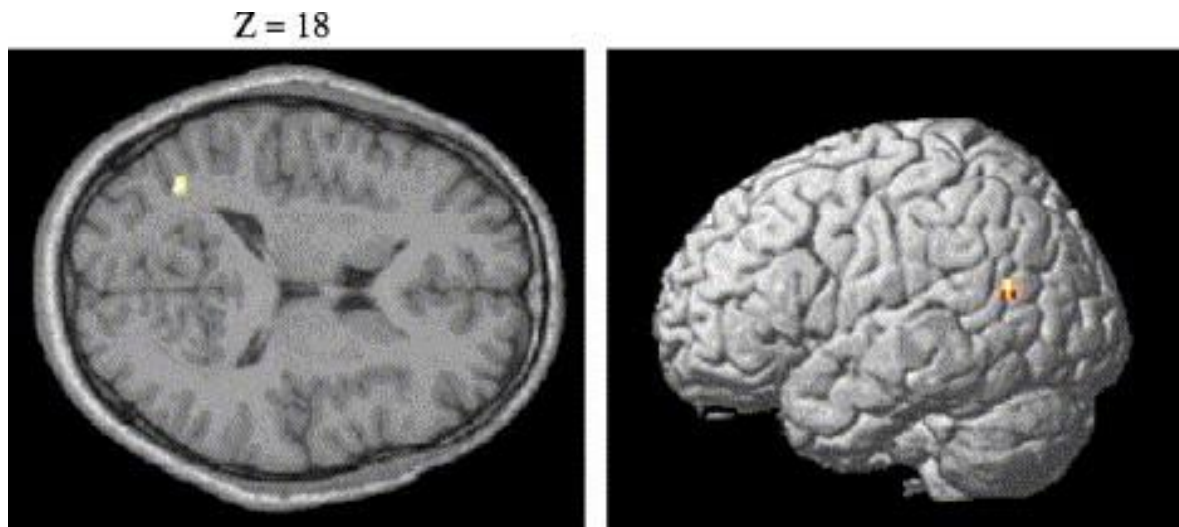


Figure 5.3. A significant difference before and after learning phonograms¹. The left image shows the axial image ($Z = 18$) and the right one shows the rendering image on the lateral surface of the left hemisphere. Displayed voxels are associated with z-scores greater than 3.09 ($P < 0.05$ corrected).

¹ Callan, Callan, and Masaki 2005:558, figure 3.

Table 5.1

Psychological interactions of BA 37¹

A significant functional connectivity for native compared with nonnative in pre-test

Region	Coordinates	Laterality	Activation maxima	
	x, y, z (mm)		Z score	Cluster size
Anterior cingulate BA24	16, -9, 45	R	4.11	39
Precentral gyrus BA6	46, -8, 34	R	3.48	36

Brain activation for $p < 0.001$ (uncorrected) with a $p < 0.05$ (uncorrected) for cluster extent. R denotes right; L denotes left.

Table 5.2

Psychological interactions of BA 39²

A significant functional connectivity for trained nonnative phonograms compared with ones before training.

Region	Coordinates	Laterality	Activation maxima	
	x, y, z (mm)		Z score	Cluster size
Caudate nucleus	-18, 25, -1	L	4.00	36
Anterior cingulate BA24	-14, 24, 14	L	3.92	22
Orbital gyrus BA11	-22, 36, -19	L	3.73	22
Superior temporal gyrus BA22	-42, -31, -5	L	3.73*	65
Occipital gyrus BA18	-12, 98, 12	L	3.13**	25

Brain activation for $p < 0.001$ (uncorrected) with a $p < 0.05$ (uncorrected) for cluster extent. R denotes right; L denotes left.

*Small-volume corrected $p = 0.006$.

**Small-volume-corrected $p = 0.027$.

¹ Callan, Callan, and Masaki 2005:558, table 1.

² Callan, Callan, and Masaki 2005:558, table 2.

5.4 Discussion

Neural activity associated with letter perception was investigated in this study. Japanese subjects who did not have any knowledge about Thai and Korean phonograms were trained on the phonograms of one of the non-native languages. Neural activity during a two-back identification task pre- and post-training was recorded. Our main interest was determining brain regions involved with visual form processing and brain regions involved with character-to-phonological conversion processing for letter perception. Letter-specific visual form processing was assessed by comparing native (Japanese) with non-native (Thai and Korean) conditions. The comparison did not reveal any significant differential brain activity. Left fusiform gyrus (BA37) failed to show differential brain activity between native and non-native conditions even when using the region of interest analyses without correction ($P > 0.05$ uncorrected). These results do not support the hypothesis that left fusiform gyrus is functionally tuned for visual form processing of letters. The result of the PPI analysis indicated that left fusiform gyrus is more strongly connected to right anterior cingulate (BA24) and right precentral gyrus (BA6) for the native condition compared to nonnative condition. In general, those areas are not considered as functional foci for the language processing. The anterior cingulate is often reported for the working memory task and the BA6 is known to be involved in motor functions. Subjects used their left hands for the button responses and right BA6 has an important role for controlling their left hands. Therefore, the enhanced functional connectivity in right BA6 may be related to the button press responses rather than visual form processing of native letters.

Character-to-phonological conversion processing was assessed by comparing post- with pre-tests for each language condition. After

training the pronunciation of non-native phonograms, left angular gyrus (BA39) was activated more during the two-back identification task. This neural change was not due to practice effects because the same change was not observed in native nor non-trained non-native conditions. It is also unlikely that the change in activity is a result of greater visual familiarity with the letters due to greater exposure as a result of training because no difference in activity was found between the very familiar native letters and the non-familiar non-native letters before training. The result of the PPI analysis indicated that left angular gyrus is more strongly connected to left caudate nucleus, left anterior cingulate (BA24), left orbital gyrus (BA11), left superior temporal gyrus, and left primary visual cortex (BA18). The enhanced integration of left angular gyrus with primary visual cortex and with superior temporal cortex supported the Wernicke–Geschwind model (Geschwind, 1965) for reading.

5.4.1 Letter-specific visual form processing

Although many previous functional imaging studies reported that letter-specific visual form processing is located in left fusiform gyrus (BA37) (Cohen et al., 2000; 2002; Polk & Farah, 1998; Polk et al., 2002; Puce et al., 1996; Tarkiainen et al., 2002), it was not observed in this study. These results are consistent with those of Joseph et al. (2003) and Price and Friston (1997) suggesting that both letters as well as objects (in this case meaningless symbols) can activate left fusiform gyrus. In our study, left fusiform gyrus was activated by both the native and non-native languages with no significant difference with respect to the degree between them. The results of our study, however, are in contrast to those of Flowers et al. (2004) in which fusiform gyrus showed greater activity to the same stimuli when they were attended to

as letters than as symbols or colors. As Flowers et al. (2004) stated; the enhanced activity in left fusiform gyrus may be caused by attending to linguistic aspects of stimuli (single letters), but not by merely perceiving the letter-specific visual form. The tactile activation of left fusiform gyrus by Braille reading in blind subjects also indicates that the activity is not specific to a visual input but more likely reflects processes involved with associating a sensory input (can be visual or tactile) with linguistic knowledge (Buchel, Price, & Friston, 1998). Many studies reported left fusiform gyrus using passive viewing tasks (e.g., passive viewing of letter strings versus face). Passive viewing has less control over the subjects mental processing during the task, so subjects probably were doing more linguistically oriented processing in their minds while they were viewing letter strings than in other conditions. It is possible that the n-back task used in our study did not require sufficient linguistic associative processing to differentially activate left fusiform gyrus. Consistent with this hypothesis differential activity in left fusiform gyrus for letters and non-letter shapes was not reported in other fMRI studies using n-back tasks (Nystrom et al., 2000; Tagamets et al., 2000).

5.4.2 Grapheme-to-Phonological Conversion Processing

The results of our study indicate that left angular gyrus is activated more in the post-test than the pre-test for the trained nonnative condition but not for the native and the non-trained nonnative conditions (Fig. 5.3). The tasks and the visual stimuli were identical in both pre- and post-tests. The only difference we assumed was the newly acquired phonological knowledge of the letters. All the subjects learned the character-to-phonological mapping for the five non-native letters they were trained on within the short training period. The

durability of these mappings in the long-term was not investigated. The issue under study here is determining the effect the newly acquired character-to-phonological mapping has on brain processes involved with the two-back task. Because the same change was not observed in the other conditions post-relative to pre-training (native and non-trained non-native), we concluded that the change in neural activity was caused by character-to-phonological conversion processing for the newly learned phonograms. The results of the interaction analysis showing greater post- minus pre-activity for the trained non-native condition relative to the native and non-trained non-native conditions further suggest that the results are due to character-to-phonological conversion rather than from practice effects or varying physiological and neurological noise.

The PPI results showed enhanced functional connectivity with left BA39 in left caudate nucleus, left anterior cingulate, left orbital gyrus, left superior temporal gyrus (BA22), and left occipital gyrus (BA18). The activation in the caudate nucleus is reported for reading and hearing (Buchsbaum et al., 2005; Vorobyev et al., 2004). The caudate nucleus is hypothesized to have a critical role in category learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998). There are a few functional imaging studies support this hypothesis (Filoteo et al., 2005; Monchi, Petrides, Petre, Worsley, & Dagher, 2001). The enhanced connectivity with the caudate may reflect the learning of the new non-native phonogram category. The most interesting finding in the PPI analysis was the enhanced coupling of left BA22 and that of left BA18 to left BA39. This indicated that the functional connection corresponding to the Wernicke–Geschwind model (Geschwind, 1965) for reading was used more for perceiving non-native phonograms after training than before training.

The behavioral results also indicate this added new process in the

post-tests. Because of practice effects, the reaction times (RTs) in the post-tests for the native and the non-trained non-native conditions were reduced significantly, but the RTs for the nonnative trained condition were not reduced. This interaction was not observed in the accuracy measurement because it may be not as sensitive as the reaction times. We maintain that the RTs were not reduced because the learned character-to-phonological mapping (that cannot be ignored when carrying out the two-back task) added extra processing time.

Consistent with our results, Lee et al. (2003) reported left angular gyrus activation in learning of Korean words by Japanese. They acquired fMRI data on the 2nd day and the 16th day of the learning. After 15 days of learning Korean words, left angular gyrus was more activated for Korean words that subjects did not learn (i.e., they did not know the meaning of the words nor the pronunciation of the words as a whole) but did know the pronunciation of some of the phonograms of the words because the same phonograms were used in Korean words that they learned. Therefore, they concluded that left angular gyrus activation was due to newly acquired character-to-phonological conversion process for the non-native phonograms.

Our findings support the classical neurological model for reading. Since Dejerine (1891, 1892) reported left angular gyrus lesion in a man who suddenly became unable to write and read (alexia with agraphia), left angular gyrus has been considered as an important site for written language processing. Based on the lesion studies, Geschwind (1965) proposed that left angular gyrus was an area associating visual and auditory information in reading. In his model, reading is processed by the route connecting, primary visual cortex, angular gyrus, and Wernicke's area.

There are numerous reasons why many functional imaging studies do not find angular gyrus activation: The use of words instead of letters

may result in lexical-to-phonological (top-down) processing that requires less grapheme-to-phonological (bottom-up) processing; the use of orthographies with ambiguous phonological mappings may not optimally activate stimulus-driven grapheme-to-phonological conversion processing. Subtraction methodology may be sensitive to differences in task and/or stimulus features between conditions that may obscure activity reflecting grapheme-to-phonological conversion processing in angular gyrus. Although many functional imaging studies using subtraction methodology did not find left angular gyrus activity, ones using functional connectivity analysis (in agreement with our results) reported that there is a strong functional connection between left angular gyrus and occipital and temporal regions in non-impaired readers and that this strong functional connection in the left hemisphere is absent in dyslexic readers (Horwitz, Rumsey, & Donohue, 1998; Pugh et al., 2000).

We tried to overcome some of the problems associated with experiments investigating phonological aspects of letter perception but there are several issues that have to be kept in mind when interpreting our results. (1) Newly learned non-native letters may not be processed in the same manner as well-experienced native letters. As can be seen in our behavioral results, character-to-phonology association training induced extra processing time in the two-back identification task. This suggests that the character-to-phonological conversion process for newly learned phonograms is not as automatic as the one for native phonograms.

(2) It is unclear whether subjects retained long-term establishment of character-to-phonological mappings for the items that they learned during the study. Learning phonograms under the experimental conditions of this study may be very different from learning phonograms in relation to acquiring a foreign language. Subjects only

learned five phonograms and all subjects could pronounce them correctly by the end of the first session out of three training sessions. The learning of the phonograms in this study may not be long-term but should nonetheless have a strong influence on neural processes involved with character-to-phonological conversion during the post-test two-back task that took place right after the training session.

(3) It is possible that the enhanced activity in left angular gyrus may not reflect the character-to-phonological conversion process but rather “extra work load” caused by newly acquired knowledge. Indeed posterior parietal (including angular gyrus) activity has been found for verbal working memory encoding tasks (Jonides et al., 1998). Because we did not include a condition to control this possibility, we cannot exclude this hypothesis. However, we believe this hypothesis is unlikely because the reported regions associated with increasing task difficulty (occipitoparietal cortices, anterior cingulate, cerebellum, right dorsolateral prefrontal, left inferior frontal gyrus, and right posterior parietal cortex) do not include left angular gyrus (Drager et al., 2004; Gould, Brown, Owen, ffytche, & Howard, 2003). Furthermore, other brain regions besides angular gyrus that are involved with working memory (Broca’s area, premotor area, supplementary motor area, dorsolateral prefrontal cortex, supramarginal gyrus, anterior cingulate, and the cerebellum; Jonides et al., 1998; Smith & Jonides, 1998; Tsukiura et al., 2001) are not differentially activated post-relative to pre-training for the learned stimuli as one would expect if the results were due only to an “extra work load”.

5.5 Conclusions

This experiment investigated brain regions involved with visual form processing and character-to-phonological (i.e., simple

grapheme-to-phonological) conversion processing for letter perception. The results of this experiment do not support the long standing hypothesis that the left fusiform gyrus (BA37) is functionally tuned for visual form processing of letters. However, consistent with the results of clinical lesion studies as well as imaging studies investigating functional connectivity, the results of this experiment strongly support the involvement of left angular gyrus (BA39) for associating visual and auditory phonological information in letter perception. The lack of left angular gyrus activation in letter or word perception for some functional imaging studies may be due to mediation by top-down processes as well as by the nature of the orthographic system of the language under study. This study attempts to control for these possible confounds by utilizing presentation of single letters from orthographic systems that are composed of phonograms with completely regular pronunciation as well as by using a methodology employing the same task and stimuli (attempting to avoid some of the problems associated with subtraction based analysis) in which differences in neural processing can be attributed predominantly to the acquisition of character-phonology association. Given the strict criteria used to assess significance in this study it is likely that some brain activity was missed in other brain regions. However, the presence of activity in left angular gyrus strongly suggests that this region is involved with character-to-phonological conversion processing in letter perception. Neural response changes associated with the newly acquired phonological knowledge of the letters provide strong support that learning can modify neural responses to linguistic stimuli in specific brain regions.

Chapter 6

Conclusions

6.1 Summary of the dissertation

Spatial processing and learning are very important information processes in human-to-human communication. In this dissertation, we investigated neural mechanisms underlying those processes by fMRI experiments.

First, we investigated spatial processing in communication and tried to find out how the brain integrates auditory and visual spatial information in speech perception. In Chapter 2, we investigated brain regions involved with auditory spatial processing in the real world. We compared neural responses between realistic spatial (externalized) auditory stimuli and artificial (internalized) auditory stimuli. The results showed that externalized auditory stimuli enhanced neural response in the pSTG compared to internalized auditory stimuli. Because of this enhanced neural response, we could observe contralateral hemifield preference in bilateral pSTG that many previous research has failed to observe. Moreover, we found that voxels in the left pSTG were activated more by back locations than front locations. These results suggest that externalized auditory stimuli (i.e. binaural recordings) are more suitable than internalized auditory stimuli (i.e. stereo recordings) to understand auditory neural processing in the real world and that the pSTG is involved with auditory spatial processing in the real world.

In Chapter 3, we discussed the neural basis of audiovisual spatial integration in speech perception. Neural substrates of the ventriloquism effect were investigated to understand the neural mechanisms of

multimodal integration. Consistent with results in Chapter 2, our fMRI experiment using externalized auditory spatial stimuli (i.e. binaural recordings) was able to show that bilateral pSTG were more sensitive to sound sources in contra- than in ipsilateral hemifields. ROI analyses in the pSTG revealed monotonically increasing neural responses for more contralateral locations. This is the first study to show sound azimuth-related BOLD signal changes in the human auditory cortex and the results support population rate coding but not topographical place coding in the human auditory cortex. The comparison between the auditory-only conditions and the audiovisual conditions successfully demonstrated the neural basis of the ventriloquism effect as arising from attenuation of the monotonically increasing functions of sound azimuth processing in the pSTG by the capture of spatially discordant visual stimuli. These results indicate that audiovisual spatial information is integrated by attenuating neural processing of auditory localization when audio and visual signals are considered to be generated from the same speaker but their spatial information is different.

Second, we investigated learning in communication and discussed how learning affects neural responses to linguistic stimuli. In Chapter 4, we tested whether different learning levels cause different neural responses in phoneme identification. Brain activity during Japanese short/long vowel identification and consonant identification are compared between native English and native Japanese speakers. The results of this experiment show enhanced activity for non-native over native speech perception of a duration phonetic contrast in brain regions involved with auditory-articulatory mapping. Unlike studies for perception of non-native spectral contrasts (Callan et al., 2004; Callan et al., 2003; Golestani and Zatorre, 2004), differential activity in the supramarginal gyrus was not found for identification of difficult

non-native durational contrasts. On the other hand, enhanced activity in the left supramarginal gyrus was found for consonant over vowel length identification by native Japanese speakers. Enhanced activity in the supramarginal gyrus for consonant identification over vowel length identification implies that the supramarginal gyrus plays a more important role for phonetic contrasts differing in place of articulation than in vowel duration. These results support the hypothesis that neural processes used to facilitate non-native phoneme perception depend on the relative contribution of information important for articulatory planning control. Differential neural activity between native and non-native phoneme perception supports the hypothesis that different learning levels cause different neural responses in phoneme identification in specific brain regions.

In Chapter 5, we investigated neural response changes related to learning of non-native linguistic stimuli. Changes in neural activity related to learning non-native orthographic to phonological mappings was examined by comparing brain activity before and after learning of previously unknown phonograms. The left angular gyrus was activated more after training the pronunciation of non-native phonograms. The result of the functional connectivity (i.e. PPI) analysis indicated that the left angular gyrus is more strongly connected to the left caudate nucleus, left anterior cingulate, left orbital gyrus, left superior temporal gyrus, and left primary visual cortex. The enhanced integration of left angular gyrus with primary visual cortex and with superior temporal cortex support the Wernicke–Geschwind model (Geschwind, 1965) for reading. The results of this experiment strongly support the involvement of left angular gyrus in orthographic to phonological mappings and provide evidence suggesting that learning can modify neural representations of linguistic stimuli in specific brain regions.

6.2 Future directions

In this dissertation, we discussed neural mechanisms underlying spatial processing and learning in human-to-human communication. Our studies added new knowledge in those areas. Additional future research is needed to extend our research for better understanding of human-to-human communication and for assisting in the development of new technologies for communication devices. There are many research topics we need to consider. We will list three research themes that we are planning to do in the near future.

The first theme is related to spatial processing and to obtain the whole response function in the right pSTG that cover 360° horizontal azimuth. Because we obtained BOLD response function only in a limited range (between -30° and 30°) in our previous study, in the future, we need to collect data to cover all 360° horizontal azimuth and examine whether the human auditory response function is similar to the monkey auditory response function reported by Werner-Reiss and Groh (2008). The second theme that needs to be investigated is concerned with the aging influence on audiovisual spatial processing. Aging can deteriorate both audition and vision. Understanding how deteriorations of auditory and vision affect audiovisual spatial processing may provide insights to improve spatial processing in aged people. The third theme to be investigated in future research is to determine individual differences in brain activity underlying language learning. Correlations of brain activity and learning performance have been reported by several studies (e.g. D. E. Callan et al., 2003). If we can classify individuals in several subgroups by using their brain activity pattern, we may be able to develop efficient teaching methods suited for each subgroup.

REFERENCES

- Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proc Natl Acad Sci U S A*, 98(21), 12301-12306. doi: 10.1073/pnas.211209098
- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol*, 14(3), 257-262. doi: 10.1016/j.cub.2004.01.029 S0960982204000430 [pii]
- Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb Cortex*, 9(5), 415-430.
- Altmann, C. F., Bledowski, C., Wibral, M., & Kaiser, J. (2007). Processing of location and pattern changes of natural sounds in the human auditory cortex. *Neuroimage*, 35(3), 1192-1200. doi: S1053-8119(07)00036-5 [pii] 10.1016/j.neuroimage.2007.01.007
- Arnott, S. R., & Alain, C. (2011). The auditory dorsal pathway: orienting vision. *Neurosci Biobehav Rev*, 35(10), 2162-2173. doi: 10.1016/j.neubiorev.2011.04.005
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychol Rev*, 105(3), 442-481.
- Baker, S. C., Frith, C. D., Frackowiak, R. S., & Dolan, R. J. (1996). Active representation of shape and spatial location in man. *Cereb Cortex*, 6(4), 612-619.
- Bates, E., Burani, C., D'Amico, S., & Barca, L. (2001). Word reading and picture naming in Italian. *Mem Cognit*, 29(7), 986-999.
- Baumgart, F., Gaschler-Markefski, B., Woldorff, M. G., Heinze, H. J., & Scheich, H. (1999). A movement-sensitive area in auditory cortex. *Nature*, 400(6746), 724-726. doi: 10.1038/23385

- Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, 41(5), 809-823. doi: S0896627304000704 [pii]
- Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, 34(1), 149-159. doi: S0896627302006426 [pii]
- Beauregard, M., Chertkow, H., Bub, D., Murtha, S., Dixon, R., & Evans, A. (1997). The neural substrate for concrete, abstract, and emotional word lexica a positron emission tomography study. *J Cogn Neurosci*, 9(4), 441-461. doi: 10.1162/jocn.1997.9.4.441
- Blauert, J. (1997). *Spatial Hearing*. Cambridge, MA. (Revised Edition): MIT Press.
- Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H. J., & Hillyard, S. A. (2007). Neural basis of the ventriloquist illusion. *Curr Biol*, 17(19), 1697-1703. doi: S0960-9822(07)01906-9 [pii] 10.1016/j.cub.2007.08.050
- Bookheimer, S. Y., Zeffiro, I. A., Blaxton, T., Gaillard, W., & Theodore, W. (1995). Regional cerebral blood flow during object naming and word reading. *Human Brain Mapping*, 3(2), 93-106.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J Acoust Soc Am*, 101(4), 2299-2310.
- Broca, P. (1861). Perte de la parole, ramollissement chronique et destruction partielle du lobe antérieur gauche du cerveau. *Bulletins de la Societe d'anthropologie, Ire serie*(2), 235-238.
- Bronkhorst, A. W., & Houtgast, T. (1999). Auditory distance perception in rooms. *Nature*, 397(6719), 517-520. doi: 10.1038/17374
- Brugge, J. F., Reale, R. A., & Hind, J. E. (1996). The structure of spatial

- receptive fields of neurons in primary auditory cortex of the cat. *Journal of Neuroscience*, 16(14), 4420-4437.
- Brunetti, M., Belardinelli, P., Caulo, M., Del Gratta, C., Della Penna, S., Ferretti, A., . . . Romani, G. L. (2005). Human brain activation during passive listening to sounds from different locations: an fMRI and MEG study. *Hum Brain Mapp*, 26(4), 251-261. doi: 10.1002/hbm.20164
- Brunswick, N., McCrory, E., Price, C. J., Frith, C. D., & Frith, U. (1999). Explicit and implicit processing of words and pseudowords by adult developmental dyslexics: A search for Wernicke's Wortschatz? *Brain*, 122 (Pt 10), 1901-1917.
- Buchel, C., Price, C., & Friston, K. (1998). A multimodal language region in the ventral visual pathway. *Nature*, 394(6690), 274-277. doi: 10.1038/28389
- Buchsbaum, B. R., Olsen, R. K., Koch, P. F., Kohn, P., Kippenhan, J. S., & Berman, K. F. (2005). Reading, hearing, and the planum temporale. *Neuroimage*, 24(2), 444-454. doi: 10.1016/j.neuroimage.2004.08.025
- Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron*, 35(4), 625-641. doi: Doi 10.1016/S0896-6273(02)00830-9
- Burke, K. A., Letsos, A., & Butler, R. A. (1994). Asymmetric performances in binaural localization of sound in space. *Neuropsychologia*, 32(11), 1409-1417. doi: 0028-3932(94)00074-3 [pii]
- Bushara, K. O., Weeks, R. A., Ishii, K., Catalan, M. J., Tian, B., Rauschecker, J. P., & Hallett, M. (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nat Neurosci*, 2(8), 759-766. doi: 10.1038/11239
- Callan, A., Callan, D., & Ando, H. (2015). An fMRI Study of the

- Ventriloquism Effect. *Cereb Cortex*. doi: 10.1093/cercor/bhu306
- Callan, A., Callan, D. E., & Ando, H. (2013). Neural correlates of sound externalization. *Neuroimage*, 66C, 22-27. doi: S1053-8119(12)01060-9 [pii] 10.1016/j.neuroimage.2012.10.057
- Callan, A. M., Callan, D. E., & Masaki, S. (2005). When meaningless symbols become letters: neural activity change in learning new phonograms. *Neuroimage*, 28(3), 553-562. doi: 10.1016/j.neuroimage.2005.06.031
- Callan, A. M., Callan, D. E., Tajima, K., & Akahane-Yamada, R. (2006). Neural processes involved with perception of non-native durational contrasts. *Neuroreport*, 17(12), 1353-1357. doi: 10.1097/01.wnr.0000224774.66904.29 00001756-200608210-00025 [pii]
- Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage*, 22(3), 1182-1194. doi: 10.1016/j.neuroimage.2004.03.006 S1053811904001508 [pii]
- Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage*, 19(1), 113-124.
- Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J. P., Berry, I., Nespoulous, J. L., & Chollet, F. (1999). Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *Neuroimage*, 9(1), 135-144. doi: 10.1006/nimg.1998.0389

- Cherry, C. (1990). *On Human Communication, third edition*. Cambridge, MA.: MIT Press.
- Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: the case for the visual word form area. *Neuroimage*, 22(1), 466-476. doi: 10.1016/j.neuroimage.2003.12.049
- Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., & Michel, F. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123 (Pt 2), 291-307.
- Cohen, L., Lehericy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, 125(Pt 5), 1054-1069.
- Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1991). Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. *J Neurosci*, 11(8), 2383-2402.
- Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1996). Object and spatial visual working memory activate separate neural systems in human cortex. *Cereb Cortex*, 6(1), 39-49.
- Damasio, A. R., & Damasio, H. (1983). The anatomic basis of pure alexia. *Neurology*, 33(12), 1573-1583.
- Dejerine, J. (1891). Sur un cas de cécité verbale avec agraphie, suivi d'autopsie. *C. R. Soc. Biol.*, 3, 197-201.
- Dejerine, J. (1892). Contribution à l'étude anatomo-pathologique et clinique des différentes variétés de cécité verbale. *C. R. Soc. Biol.*, 4.
- Demonet, J. F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J. L., Wise, R., . . . Frackowiak, R. (1992). The anatomy of phonological

- and semantic processing in normal subjects. *Brain*, 115 (Pt 6), 1753-1768.
- Deouell, L. Y., Heller, A. S., Malach, R., D'Esposito, M., & Knight, R. T. (2007). Cerebral responses to change in spatial location of unattended sounds. *Neuron*, 55(6), 985-996. doi: S0896-6273(07)00648-4 [pii] 10.1016/j.neuron.2007.08.019
- Drager, B., Jansen, A., Bruchmann, S., Forster, A. F., Pleger, B., Zwitserlood, P., & Knecht, S. (2004). How does the brain accommodate to increased task difficulty in word finding? A functional MRI study. *Neuroimage*, 23(3), 1152-1160. doi: 10.1016/j.neuroimage.2004.07.005
- Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Jr., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, 92(1-2), 145-177. doi: 10.1016/j.cognition.2003.11.002
- Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, 12(10), 388-396. doi: DOI 10.1016/j.tics.2008.07.004
- Erickson, C. A., Jagadeesh, B., & Desimone, R. (2000). Clustering of perirhinal neurons with similar properties following visual experience in adult monkeys. *Nat Neurosci*, 3(11), 1143-1148. doi: 10.1038/80664
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur J Neurosci*, 15(2), 399-402.
- Fant, G. (1960). *Acoustic theory of speech production*. Mouton: The Hague.
- Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proc Natl Acad Sci U S A*, 108(39), 16428-16433. doi:

10.1073/pnas.1112937108

- Fiez, J. A., Balota, D. A., Raichle, M. E., & Petersen, S. E. (1999). Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading. *Neuron*, 24(1), 205-218.
- Filoteo, J. V., Maddox, W. T., Simmons, A. N., Ing, A. D., Cagigas, X. E., Matthews, S., & Paulus, M. P. (2005). Cortical and subcortical brain regions involved in rule-based category learning. *Neuroreport*, 16(2), 111-115.
- Flowers, D. L., Jones, K., Noble, K., VanMeter, J., Zeffiro, T. A., Wood, F. B., & Eden, G. F. (2004). Attention to single letters activates left extrastriate cortex. *Neuroimage*, 21(3), 829-839. doi: 10.1016/j.neuroimage.2003.10.002
- Fridriksson, J., Fillmore, P., Guo, D., & Rorden, C. (2014). Chronic Broca's Aphasia Is Caused by Damage to Broca's and Wernicke's Areas. *Cereb Cortex*. doi: 10.1093/cercor/bhu152
- Frissen, I., Vroomen, J., & de Gelder, B. (2012). The aftereffects of ventriloquism: the time course of the visual recalibration of auditory localization. *Seeing Perceiving*, 25(1), 1-14. doi: 10.1163/187847611X620883
- Frissen, I., Vroomen, J., de Gelder, B., & Bertelson, P. (2005). The aftereffects of ventriloquism: generalization across sound-frequencies. *Acta Psychol (Amst)*, 118(1-2), 93-100. doi: 10.1016/j.actpsy.2004.10.004
- Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*, 6(3), 218-229. doi: 10.1006/nimg.1997.0291
- Fujimaki, N., Miyauchi, S., Putz, B., Sasaki, Y., Takino, R., Sakai, K., & Tamada, T. (1999). Functional magnetic resonance imaging of neural activity related to orthographic, phonological, and lexico-semantic

- judgments of visually presented characters and words. *Hum Brain Mapp*, 8(1), 44-59.
- Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Tong, Y., & Li, X. (2002). A cross-linguistic fMRI study of spectral and temporal cues underlying phonological processing. *J Cogn Neurosci*, 14(7), 1076-1087. doi: 10.1162/089892902320474526
- Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Tong, Y., & Lurito, J. (2002). Neural circuitry underlying perception of duration depends on language experience. *Brain Lang*, 83(2), 268-290.
- Gelfand, J. R., & Bookheimer, S. Y. (2003). Dissociating neural mechanisms of temporal sequencing and processing phonemes. *Neuron*, 38(5), 831-842.
- Geschwind, N. (1965). Disconnexion syndromes in animals and man. I. *Brain*, 88(2), 237-294.
- Getzmann, S., & Lewald, J. (2010a). Effects of natural versus artificial spatial cues on electrophysiological correlates of auditory motion. *Hear Res*, 259(1-2), 44-54. doi: S0378-5955(09)00245-7 [pii] 10.1016/j.heares.2009.09.021
- Getzmann, S., & Lewald, J. (2010b). Shared cortical systems for processing of horizontal and vertical sound motion. *J Neurophysiol*, 103(4), 1896-1904. doi: 00333.2009 [pii] 10.1152/jn.00333.2009
- Gitelman, D. R., Penny, W. D., Ashburner, J., & Friston, K. J. (2003). Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. *Neuroimage*, 19(1), 200-207.
- Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: reallocation of neural substrates. *Neuroimage*, 21(2), 494-506. doi: 10.1016/j.neuroimage.2003.09.071
- Goodman, D. F., & Brette, R. (2010). Spike-timing-based computation in

- sound localization. *PLoS Comput Biol*, 6(11), e1000993. doi: 10.1371/journal.pcbi.1000993
- Gould, R. L., Brown, R. G., Owen, A. M., ffytche, D. H., & Howard, R. J. (2003). fMRI BOLD response to increasing task difficulty during successful paired associates learning. *Neuroimage*, 20(2), 1006-1019. doi: 10.1016/S1053-8119(03)00365-3
- Granacher, R. P. J. (2015). *Traumatic Brain Injury*. Boca Raton, FL: CRC Press.
- Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends Neurosci*, 25(7), 348-353. doi: S0166223602021914 [pii]
- Grodd, W., Hulsman, E., Lotze, M., Wildgruber, D., & Erb, M. (2001). Sensorimotor mapping of the human cerebellum: fMRI evidence of somatotopic organization. *Hum Brain Mapp*, 13(2), 55-73.
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang*, 96(3), 280-301. doi: S0093-934X(05)00115-X [pii] 10.1016/j.bandl.2005.06.001
- Guenther, F. H., & Perkell, J. S. (2004). A neural model of speech production and its application to studies of the role of auditory feedback in speech. In B. Maassen, K. R., H. Peters, P. Van Lieshout, & W. Hulstijn (Eds.), *Speech Motor Control in Normal and Disordered Speech*. (pp. 29-49). Oxford: Oxford University Press.
- Gusnard, D. A., Raichle, M. E., & Raichle, M. E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nat Rev Neurosci*, 2(10), 685-694. doi: 10.1038/35094500
- Hartmann, W. M., & Wittenberg, A. (1996). On the externalization of sound images. *J Acoust Soc Am*, 99(6), 3678-3688.
- Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., & Grady, C. L. (1994). The functional organization of human

- extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *J Neurosci*, 14(11 Pt 1), 6336-6353.
- Henderson, V. W. (1986). Anatomy of posterior pathways in reading: a reassessment. *Brain Lang*, 29(1), 119-133.
- Horwitz, B., Rumsey, J. M., & Donohue, B. C. (1998). Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proc Natl Acad Sci U S A*, 95(15), 8939-8944.
- Howard, D., Patterson, K., Wise, R., Brown, W. D., Friston, K., Weiller, C., & Frackowiak, R. (1992). The cortical localization of the lexicons. Positron emission tomography evidence. *Brain*, 115 (Pt 6), 1769-1782.
- Hudson, P. T. W., & Bergman, M. W. (1984). Lexical knowledge in word recognition: word length and word frequency in naming and lexical decision tasks. *J. Mem. Lang.*, 24, 46-58.
- Indefrey, P., Kleinschmidt, A., Merboldt, K. D., Kruger, G., Brown, C., Hagoort, P., & Frahm, J. (1997). Equivalent responses to lexical and nonlexical visual stimuli in occipital cortex: a functional magnetic resonance imaging study. *Neuroimage*, 5(1), 78-81. doi: 10.1006/nimg.1996.0232
- Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Percept Mot Skills*, 37(3), 967-979.
- Jeffress, L. A. (1948). A place theory of sound localization. *J Comp Physiol Psychol*, 41(1), 35-39.
- Jonides, J., Schumacher, E. H., Smith, E. E., Koeppe, R. A., Awh, E., Reuter-Lorenz, P. A., . . . Willis, C. R. (1998). The role of parietal cortex in verbal working memory. *J Neurosci*, 18(13), 5026-5034.
- Joseph, J. E., Gathers, A. D., & Piper, G. A. (2003). Shared and dissociated cortical regions for object and letter processing. *Brain Res Cogn Brain Res*, 17(1), 56-67.

- Joubert, S., Beauregard, M., Walter, N., Bourgouin, P., Beaudoin, G., Leroux, J. M., . . . Lecours, A. R. (2004). Neural correlates of lexical and sublexical processes in reading. *Brain Lang*, 89(1), 9-20. doi: 10.1016/S0093-934X(03)00403-6
- Kaiser, J., Lutzenberger, W., Preissl, H., Ackermann, H., & Birbaumer, N. (2000). Right-hemisphere dominance for the processing of sound-source lateralization. *J Neurosci*, 20(17), 6631-6639. doi: 20/17/6631 [pii]
- Knudsen, E. I., & Konishi, M. (1978). A neural map of auditory space in the owl. *Science*, 200(4343), 795-797.
- Konishi, M. (2003). Coding of auditory space. *Annu Rev Neurosci*, 26, 31-55. doi: 10.1146/annurev.neuro.26.041002.131123
- Krumbholz, K., Nobis, E. A., Weatheritt, R. J., & Fink, G. R. (2009). Executive control of spatial attention shifts in the auditory compared to the visual modality. *Hum Brain Mapp*, 30(5), 1457-1469. doi: 10.1002/hbm.20615
- Krumbholz, K., Schonwiesner, M., Rubsamen, R., Zilles, K., Fink, G. R., & von Cramon, D. Y. (2005). Hierarchical processing of sound location and motion in the human brainstem and planum temporale. *Eur J Neurosci*, 21(1), 230-238. doi: EJN3836 [pii] 10.1111/j.1460-9568.2004.03836.x
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos Trans R Soc Lond B Biol Sci*, 363(1493), 979-1000. doi: 10.1098/rstb.2007.2154
- Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, 218(4577), 1138-1141.
- Kumpik, D. P., Kacelnik, O., & King, A. J. (2010). Adaptive reweighting of auditory localization cues in response to chronic unilateral

- earplugging in humans. *J Neurosci*, 30(14), 4883-4894. doi: 30/14/4883 [pii] 10.1523/JNEUROSCI.5488-09.2010
- Kuriki, S., Hirata, Y., Fujimaki, N., & Kobayashi, T. (1996). Magnetoencephalographic study on the cerebral neural activities related to the processing of visually presented characters. *Brain Res Cogn Brain Res*, 4(3), 185-199.
- Law, I., Kanna, I., Fujita, H., Lassen, N. A., Miura, S., & Uemura, K. (1991). Left supramarginal/angular gyri activation during reading of syllabograms in the Japanese language. *J. Neurolinguist*, 6, 243-251.
- Lee, H. S., Fujii, T., Okuda, J., Tsukiura, T., Umetsu, A., Suzuki, M., . . . Yamadori, A. (2003). Changes in brain activation patterns associated with learning of Korean words by Japanese: an fMRI study. *Neuroimage*, 20(1), 1-11.
- Leung, H. C., Gore, J. C., & Goldman-Rakic, P. S. (2002). Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. *Journal of Cognitive Neuroscience*, 14(4), 659-671. doi: Doi 10.1162/08989290260045882
- Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., & DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cereb Cortex*, 14(9), 1008-1021. doi: 10.1093/cercor/bhh061 bhh061 [pii]
- Linzenbold, W., & Himmelbach, M. (2012). Signals from the deep: reach-related activity in the human superior colliculus. *J Neurosci*, 32(40), 13881-13888. doi: 10.1523/JNEUROSCI.0619-12.2012
- Logothetis, N. K. (2002). The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philos Trans R Soc Lond B Biol Sci*, 357(1424), 1003-1037. doi: 10.1098/rstb.2002.1114
- Luria, A. R. (1973). *The Working Brain: An Introduction to*

- Neuropsychology*. New York: Basic Books, Inc.
- Magezi, D. A., & Krumbholz, K. (2010). Evidence for Opponent-Channel Coding of Interaural Time Differences in Human Auditory Cortex. *Journal of Neurophysiology*, 104(4), 1997-2007. doi: DOI 10.1152/jn.00424.2009
- Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mathiak, K., Hertrich, I., Grodd, W., & Ackermann, H. (2002). Cerebellum and speech perception: a functional magnetic resonance imaging study. *J Cogn Neurosci*, 14(6), 902-912. doi: 10.1162/089892902760191126
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746-748.
- Mershon, D. H., & Bowers, J. N. (1979). Absolute and relative cues for the auditory perception of egocentric distance. *Perception*, 8(3), 311-322.
- Middlebrooks, J. C., & Knudsen, E. I. (1984). A neural code for auditory space in the cat's superior colliculus. *J Neurosci*, 4(10), 2621-2634.
- Middlebrooks, J. C., Xu, L., Eddins, A. C., & Green, D. M. (1998). Codes for sound-source location in nontopographic auditory cortex. *J Neurophysiol*, 80(2), 863-881.
- Minagawa-Kawai, Y., Mori, K., Sato, Y., & Koizumi, T. (2004). Differential cortical responses in second language learners to different vowel contrasts. *Neuroreport*, 15(5), 899-903.
- Miyawaki, K., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Percept Psychophys*, 18(5), 331-340.
- Monchi, O., Petrides, M., Petre, V., Worsley, K., & Dagher, A. (2001). Wisconsin Card Sorting revisited: distinct neural circuits

- participating in different stages of the task identified by event-related functional magnetic resonance imaging. *J Neurosci*, 21(19), 7733-7741.
- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. *Cogn Affect Behav Neurosci*, 5(3), 263-281.
- Nystrom, L. E., Braver, T. S., Sabb, F. W., Delgado, M. R., Noll, D. C., & Cohen, J. D. (2000). Working memory for letters, shapes, and locations: fMRI evidence against stimulus-based regional organization in human prefrontal cortex. *Neuroimage*, 11(5 Pt 1), 424-446. doi: 10.1006/nimg.2000.0572
- Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A*, 87(24), 9868-9872.
- Palmer, A. R., & King, A. J. (1982). The representation of auditory space in the mammalian superior colliculus. *Nature*, 299(5880), 248-249.
- Palomaki, K. J., Tiitinen, H., Makinen, V., May, P. J., & Alku, P. (2005). Spatial processing in human auditory cortex: the effects of 3D, ITD, and ILD stimulation techniques. *Brain Res Cogn Brain Res*, 24(3), 364-379. doi: S0926-6410(05)00048-0 [pii] 10.1016/j.cogbrainres.2005.02.013
- Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362(6418), 342-345. doi: 10.1038/362342a0
- Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., . . . Frith, U. (2000). A cultural effect on brain function. *Nat Neurosci*, 3(1), 91-96. doi: 10.1038/71163
- Pavani, F., Macaluso, E., Warren, J. D., Driver, J., & Griffiths, T. D. (2002). A common cortical substrate activated by horizontal and

- vertical sound movement in the human brain. *Curr Biol*, 12(18), 1584-1590. doi: S0960982202011430 [pii]
- Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, 331(6157), 585-589. doi: 10.1038/331585a0
- Plenge, G. (1974). On the differences between localization and lateralization. *J Acoust Soc Am*, 56(3), 944-951.
- Poeppel, D. (1996). A critical review of PET studies of phonological processing. *Brain Lang*, 55(3), 317-351; discussion 352-385. doi: 10.1006/brln.1996.0108
- Poldrack, R. A. (2000). Imaging brain plasticity: conceptual and methodological issues--a theoretical review. *Neuroimage*, 12(1), 1-13. doi: 10.1006/nimg.2000.0596
- S1053-8119(00)90596-2 [pii]
- Polk, T. A., & Farah, M. J. (1998). The neural development and organization of letter recognition: evidence from functional neuroimaging, computational modeling, and behavioral studies. *Proc Natl Acad Sci U S A*, 95(3), 847-852.
- Polk, T. A., Stallcup, M., Aguirre, G. K., Alsop, D. C., D'Esposito, M., Detre, J. A., & Farah, M. J. (2002). Neural specialization for letter recognition. *J Cogn Neurosci*, 14(2), 145-159. doi: 10.1162/089892902317236803
- Pralong, D. (1996). The role of individualized headphone calibration for the generation of high fidelity virtual auditory space. *J Acoust Soc Am*, 100(6), 3785-3793.
- Price, C. J. (2000). The anatomy of language: contributions from functional neuroimaging. *J Anat*, 197 Pt 3, 335-359.
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading.

- Neuroimage*, 62(2), 816-847. doi: S1053-8119(12)00470-3 [pii]
10.1016/j.neuroimage.2012.04.062
- Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *Neuroimage*, 19(3), 473-481.
- Price, C. J., & Devlin, J. T. (2004). The pro and cons of labelling a left occipitotemporal region: "the visual word form area". *Neuroimage*, 22(1), 477-479. doi: 10.1016/j.neuroimage.2004.01.018
- Price, C. J., & Friston, K. J. (1997). Cognitive conjunction: a new approach to brain activation experiments. *Neuroimage*, 5(4 Pt 1), 261-270. doi: 10.1006/nimg.1997.0269
- Price, C. J., Moore, C. J., Humphreys, G. W., & Wise, R. J. (1997). Segregating Semantic from Phonological Processes during Reading. *J Cogn Neurosci*, 9(6), 727-733. doi: 10.1162/jocn.1997.9.6.727
- Price, C. J., Wise, R. J., Watson, J. D., Patterson, K., Howard, D., & Frackowiak, R. S. (1994). Brain activity during reading. The effects of exposure duration and task. *Brain*, 117 (Pt 6), 1255-1269.
- Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J Neurosci*, 16(16), 5205-5215.
- Pugh, K. R., Mencl, W. E., Shaywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Constable, R. T., . . . Gore, J. C. (2000). The angular gyrus in developmental dyslexia: task-specific differences in functional connectivity within posterior cortex. *Psychol Sci*, 11(1), 51-56.
- Pugh, K. R., Shaywitz, B. A., Shaywitz, S. E., Constable, R. T., Skudlarski, P., Fulbright, R. K., . . . Gore, J. C. (1996). Cerebral organization of component processes in reading. *Brain*, 119 (Pt 4), 1221-1238.
- Rajan, R., Aitkin, L. M., Irvine, D. R. F., & Mckay, J. (1990). Azimuthal Sensitivity of Neurons in Primary Auditory-Cortex of Cats .1. Types

- of Sensitivity and the Effects of Variations in Stimulus Parameters. *Journal of Neurophysiology*, 64(3), 872-887.
- Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Curr Opin Neurobiol*, 8(4), 516-521.
- Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci U S A*, 97(22), 11800-11806. doi: 10.1073/pnas.97.22.11800
- Rauschecker, J. P., Tian, B., & Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*, 268(5207), 111-114.
- Recanzone, G. H. (1998). Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proc Natl Acad Sci U S A*, 95(3), 869-875.
- Renier, L. A., Anurova, I., De Volder, A. G., Carlson, S., VanMeter, J., & Rauschecker, J. P. (2010). Preserved Functional Specialization for Spatial Processing in the Middle Occipital Gyrus of the Early Blind. *Neuron*, 68(1), 138-148. doi: DOI 10.1016/j.neuron.2010.09.021
- Roby-Brami, A., Hermsdorfer, J., Roy, A. C., & Jacobs, S. (2012). A neuropsychological perspective on the link between language and praxis in modern humans. *Philos Trans R Soc Lond B Biol Sci*, 367(1585), 144-160. doi: 10.1098/rstb.2011.0122
- Rorden, C., & Karnath, H. O. (2004). Using human brain lesions to infer function: a relic from a past era in the fMRI age? *Nat Rev Neurosci*, 5(10), 813-819. doi: 10.1038/nrn1521
- Rorden, C., Karnath, H. O., & Bonilha, L. (2007). Improving lesion-symptom mapping. *J Cogn Neurosci*, 19(7), 1081-1088. doi: 10.1162/jocn.2007.19.7.1081
- Salminen, N. H., May, P. J., Alku, P., & Tiitinen, H. (2009). A population rate code of auditory space in the human cortex. *PLoS One*, 4(10), e7600. doi: 10.1371/journal.pone.0007600
- Salminen, N. H., Tiitinen, H., Miettinen, I., Alku, P., & May, P. J. (2010).

- Asymmetrical representation of auditory space in human cortex. *Brain Res*, 1306, 93-99. doi: S0006-8993(09)02085-X [pii] 10.1016/j.brainres.2009.09.095
- Schonwiesner, M., Krumbholz, K., Rubsamen, R., Fink, G. R., & von Cramon, D. Y. (2007). Hemispheric asymmetry for auditory processing in the human auditory brain stem, thalamus, and cortex. *Cereb Cortex*, 17(2), 492-499. doi: bhj165 [pii] 10.1093/cercor/bhj165
- Schorr, E. A., Fox, N. A., van Wassenhove, V., & Knudsen, E. I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *Proc Natl Acad Sci U S A*, 102(51), 18748-18750. doi: 10.1073/pnas.0508862102
- Seifritz, E., Neuhoff, J. G., Bilecen, D., Scheffler, K., Mustovic, H., Schachinger, H., . . . Di Salle, F. (2002). Neural processing of auditory looming in the human brain. *Curr Biol*, 12(24), 2147-2151. doi: S0960982202013568 [pii]
- Sergent, J., Zuck, E., Levesque, M., & MacDonald, B. (1992). Positron emission tomography study of letter and object processing: empirical findings and methodological considerations. *Cereb Cortex*, 2(1), 68-80.
- Shinn-Cunningham, B. G., Kopco, N., & Martin, T. J. (2005). Localizing nearby sound sources in a classroom: binaural room impulse responses. *J Acoust Soc Am*, 117(5), 3100-3115.
- Small, S. L., Noll, D. C., Perfetti, C. A., Hlustik, P., Wellington, R., & Schneider, W. (1996). Localizing the lexicon for reading aloud: replication of a PET study using fMRI. *Neuroreport*, 7(4), 961-965.
- Smith, E. E., & Jonides, J. (1998). Neuroimaging analyses of human working memory. *Proc Natl Acad Sci U S A*, 95(20), 12061-12068.
- Stecker, G. C., Harrington, I. A., & Middlebrooks, J. C. (2005). Location

- coding by opponent neural populations in the auditory cortex. *Plos Biology*, 3(3), 520-528. doi: ARTN e78 DOI 10.1371/journal.pbio.0030078
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci*, 9(4), 255-266. doi: 10.1038/nrn2331
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.*, 28, 643-662.
- Suvorov, N. F., & Gulyakov, M. B. (1988). Role of the caudate nucleus in performance of the conditioned reactions connected with sound localization. *Int J Psychophysiol*, 6(2), 91-97.
- Tagamets, M. A., Novick, J. M., Chalmers, M. L., & Friedman, R. B. (2000). A parametric approach to orthographic processing in the brain: an fMRI study. *J Cogn Neurosci*, 12(2), 281-297.
- Takahashi, T., Moiseff, A., & Konishi, M. (1984). Time and intensity cues are processed independently in the auditory system of the owl. *J Neurosci*, 4(7), 1781-1786.
- Talairach, J., & Tournoux, P. (1988). *A Co-planer Stereotaxic Atlas of the Human Brain*. Stuttgart: Thieme.
- Tarkiainen, A., Cornelissen, P. L., & Salmelin, R. (2002). Dynamics of visual feature analysis and object-level processing in face versus letter-string perception. *Brain*, 125(Pt 5), 1125-1136.
- Thurlow, W. R., & Jack, C. E. (1973). Certain determinants of the "ventriloquism effect". *Percept Mot Skills*, 36(3), 1171-1184.
- Toda, T. (2003). Acquisition of Special Morae in Japanese as a Second Language. *J Phonet Soc Jpn*, 7, 70-83.
- Tsukiura, T., Fujii, T., Takahashi, T., Xiao, R., Inase, M., Iijima, T., . . . Okuda, J. (2001). Neuroanatomical discrimination between manipulating and maintaining processes involved in verbal working memory; a functional MRI study. *Brain Res Cogn Brain Res*, 11(1),

13-21.

- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., . . . Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15(1), 273-289. doi: DOI 10.1006/nimg.2001.0978
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549-586). Cambridge, MA: MIT Press.
- Vander Wyk, B. C., Ramsay, G. J., Hudac, C. M., Jones, W., Lin, D., Klin, A., . . . Pelphrey, K. A. (2010). Cortical integration of audio-visual speech and non-speech stimuli. *Brain Cogn*, 74(2), 97-106. doi: 10.1016/j.bandc.2010.07.002
- Vorobyev, V. A., Alho, K., Medvedev, S. V., Pakhomov, S. V., Roudas, M. S., Rutkovskaya, J. M., . . . Naatanen, R. (2004). Linguistic processing in visual and modality-nonspecific brain areas: PET recordings during selective attention. *Brain Res Cogn Brain Res*, 20(2), 309-322. doi: 10.1016/j.cogbrainres.2004.03.011
- Wager, T. D., & Nichols, T. E. (2003). Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *Neuroimage*, 18(2), 293-309.
- Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., . . . Buckner, R. L. (1998). Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, 281(5380), 1188-1191.
- Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual-auditory "compellingness" in the ventriloquism effect: implications for transitivity among the spatial senses. *Percept Psychophys*, 30(6), 557-564.
- Warren, J. D., Zielinski, B. A., Green, G. G., Rauschecker, J. P., &

- Griffiths, T. D. (2002). Perception of sound-source motion by the human brain. *Neuron*, 34(1), 139-148. doi: S0896627302006372 [pii]
- Warren, J. E., Wise, R. J., & Warren, J. D. (2005). Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends Neurosci*, 28(12), 636-643. doi: 10.1016/j.tins.2005.09.010
- Watkins, K., & Paus, T. (2004). Modulation of motor excitability during speech perception: the role of Broca's area. *J Cogn Neurosci*, 16(6), 978-987. doi: 10.1162/0898929041502616
- Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41(8), 989-994.
- Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, 103(1), 147-162. doi: 10.1016/j.cognition.2006.03.006
- Werner-Reiss, U., & Groh, J. M. (2008). A rate code for sound azimuth in monkey auditory cortex: implications for human neuroimaging studies. *J Neurosci*, 28(14), 3747-3758. doi: 28/14/3747 [pii] 10.1523/JNEUROSCI.5044-07.2008
- Wernicke, C. (1874). *Der aphasische Symptomenkomplex*. Breslau, Poland: Cohen and Weigert.
- Wightman, F. L., & Kistler, D. J. (1989). Headphone simulation of free-field listening. II: Psychophysical validation. *J Acoust Soc Am*, 85(2), 868-878.
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nat Neurosci*, 7(7), 701-702. doi: 10.1038/nn1263 nn1263 [pii]
- Witten, I. B., & Knudsen, E. I. (2005). Why seeing is believing: Merging

- auditory and visual worlds. *Neuron*, 48(3), 489-496. doi: DOI 10.1016/j.neuron.2005.10.020
- Woldorff, M. G., Tempelmann, C., Fell, J., Tegeler, C., Gaschler-Markefski, B., Hinrichs, H., . . . Scheich, H. (1999). Lateralized auditory spatial perception and the contralaterality of cortical processing as studied with functional magnetic resonance imaging and magnetoencephalography. *Hum Brain Mapp*, 7(1), 49-66. doi: 10.1002/(SICI)1097-0193(1999)7:1<49::AID-HBM5>3.0.CO;2-J [pii]
- Woods, T. M., & Recanzone, G. H. (2004). Visually induced plasticity of auditory spatial perception in macaques. *Curr Biol*, 14(17), 1559-1564. doi: 10.1016/j.cub.2004.08.059 S0960982204006554 [pii]
- Wozny, D. R., & Shams, L. (2011). Recalibration of auditory space following milliseconds of cross-modal discrepancy. *J Neurosci*, 31(12), 4607-4612. doi: 10.1523/JNEUROSCI.6079-10.2011
- Wydell, T. N., Vuorinen, T., Helenius, P., & Salmelin, R. (2003). Neural correlates of letter-string length and lexicality during reading in a regular orthography. *J Cogn Neurosci*, 15(7), 1052-1062. doi: 10.1162/089892903770007434
- Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, 11(10), 946-953. doi: DOI 10.1093/cercor/11.10.946
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, 6(1), 37-46. doi: Doi 10.1016/S1364-6613(00)01816-7
- Zatorre, R. J., Bouffard, M., Ahad, P., & Belin, P. (2002). Where is 'where' in the human auditory cortex? *Nat Neurosci*, 5(9), 905-909. doi: 10.1038/nn904 nn904 [pii]

- Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256(5058), 846-849.
- Zimmer, U., Lewald, J., Erb, M., & Karnath, H. O. (2006). Processing of auditory spatial cues in human cortex: an fMRI study. *Neuropsychologia*, 44(3), 454-461. doi: S0028-3932(05)00206-X [pii] 10.1016/j.neuropsychologia.2005.05.021
- Zimmer, U., & Macaluso, E. (2005). High binaural coherence determines successful sound localization and increased activity in posterior auditory areas. *Neuron*, 47(6), 893-905. doi: S0896-6273(05)00613-6 [pii] 10.1016/j.neuron.2005.07.019