이학박사학위논문

Neural mechanism of verbal repetition:

From sounds to speech

따라말하기의 신경기제: 소리에서 언어로

2013 년 2 월

서울대학교 대학원

협동과정 인지과학 전공

유 세 진

이학박사학위논문

Neural mechanism of verbal repetition:

From sounds to speech

따라말하기의 신경기제: 소리에서 언어로

2013 년 2 월

서울대학교 대학원

협동과정 인지과학 전공

유 세 진

Neural mechanism of verbal repetition: From sounds to speech

지도교수 이 경 민

이 논문을 이학박사 학위논문으로 제출함.

2012 년 12 월

서울대학교 대학원

협동과정 인지과학 전공

유 세 진

유세진의 이학박사 학위논문을 인준함.

2012 년 12 월

위 원 장 _____ (印)

부위원장 _____ (印)

위 원 _____ (印)

위 원 _____ (印)

위 원 _____ (印)

Abstract

Neural mechanism of verbal repetition: From sounds to speech

Sejin Yoo
Interdisciplinary Program in Cognitive Science
The Graduate School
Seoul National University

Verbal repetition is one of simple and natural tasks, in which both ends of speech perception and production are recruited simultaneously, and is supposed to be a fundamental tool of language acquisition, especially in word learning. In the present study, by investigating verbal repetition, we aimed at examining (1) how speech codes are represented in human brain; (2) speech hemodynamics during listening and verbal repetition of various auditory sounds, which supports the first findings, and (3) neural circuits recruited for associating meanings with novel sounds to build speech codes.

In the first experiment, we introduced novel sounds perceived as words or pseudowords depending on the interpretation of ambiguous vowel in the stimuli. With an event-related fMRI, we found the audition-articulation interface at the Sylvian fissures and superior temporal sulci bilaterally, and more importantly, we found neural activities unique to word-perceived repetition in the left posterior middle temporal areas and those unique to pseudoword-perceived repetition in

the left inferior frontal gyrus by contrasting word-versus-pseudoword trials. These findings imply that even for acoustically identical sounds, two distinct speech codes, i.e. an articulation-based code of pseudowords and an acoustic-phonetic code of words, are differentially used for verbal repetition according to whether the speech sounds are meaningful or not.

In the second experiment, we re-examined the previous findings in the first experiment with regard to hemodynamics measured by fNIRS. We monitored the hemoglobin concentration change at inferior frontal gyri bilaterally while the subjects listened to various sounds, i.e. natural sounds, animal vocalizations, human emotional sounds, pseudowords, and words, and verbally repeated speech sounds (pseudowords and words) only. We observed oxygenated hemoglobin (O_2Hb) change at left inferior frontal gyrus was positive for both speech and nonspeech sounds, but negative at right inferior frontal gyrus. Furthermore, there was hemodynamic modulation by sound types at the IFG even in passive listening. Contrasting verbal repetition of words and pseudowords revealed that the proportion of O_2Hb change in total Hb concentration was significantly higher for pseudowords than for words, indicating that articulatory codes at the LIFG were predominant for pseudowords, not for words.

In the third experiment, we further investigated how speech sounds become meaningful, i.e. what neural mechanism supports the learning process. We designed a simple associative learning paradigm combined with fMRI. For the associative learning, some novel sounds were presented with meanings in simple stories (learned condition), while others were presented without meanings in the

same stories (unlearned condition). We contrasted verbal repetition of the novel sounds before and after the learning phase. The results revealed that unlearned sounds uniquely evoked neural activities at superior and middle frontal gyri bilaterally, whereas learned sounds uniquely evoked neural activities at superior and inferior parietal lobules as well as superior and middle frontal gyri bilaterally. A connectivity analysis using dynamic causal modeling (DCM) suggested that the dorsal fronto-parietal network might subserve as episodic buffers used for associative learning of novel sounds.

Putting all together, we found that the dorsal fronto-parietal network was recruited for associative learning that novel sounds were transformed into speech sounds, i.e. meaningful sounds. Once sounds become meaningful by learning, an acoustic-phonetic code at left middle temporal gyrus was used to represent the sounds, while meaningless sounds were temporarily maintained as an articulatory code at left inferior frontal gyrus. These findings were additionally confirmed by hemodynamics of speech processing at inferior frontal gyri, indicating that speech perception might be partly dependent on generation of speech codes for speech production.

Keywords: Verbal repetition, Word learning, fMRI, fNIRS, Speech codes, DCM

Student Number: 2009-30811

Table of Contents

List of Tables

List of Figures

Chapter 1. Theoretical Background

The heavens declare the glory of God: and the firmament sheweth his handywork.

Day unto day uttereth speech, and night unto night sheweth knowledge.

There is no speech nor language, where their voice is not heard.

Their line is gone out through all the earth, and their words to the end of the world.

Psalms 19:1-4

Speech is a kind of online mental process, in which specific information is dynamically transferred between speaker and listener via relevant transferring medium, i.e. sound waves. Here, human brain operates as an adaptive processor capable of decoding and encoding the information embedded onto sound waves for successful perception and production, respectively. Of course, it is intriguing how the information is organized in its linguistic structures and to what extent the linguistic structures are predetermined before learning process. More interesting will be, however, how speech comes out of sound waves. In other words, what is the difference between speech and sounds and how can we learn speech from sounds? In the present study, we will investigate what neural mechanisms support such processes.

1. Speech processing in human brain

Speech processing in human brain can be accounted for as information flow (Denes and Pinson, 1993). In terms of information hierarchy, it is divided into three processing levels: acoustic, physiological, and linguistic levels (see Figure 1). Linguistic information is transformed into different forms at each level and naturally conveyed from one level to another. Chomskyan linguistics (Chomsky, 2000), basically in the line of structural linguistics emerging at early twentieth century, largely revealed how information is represented and organized within many different linguistic structures. However, it is still unclear how linguistic information can be extracted from specific sounds during online speech processing. Besides, there is no complete description of how acoustic sounds become meaningful speech sounds yet, which is essential in early language acquisition, too. It is thus necessary to reveal neural processes mediating between any two adjacent levels of speech processing, i.e. acoustic-to-physiological and physiological-to-linguistic processes, and vice versa.

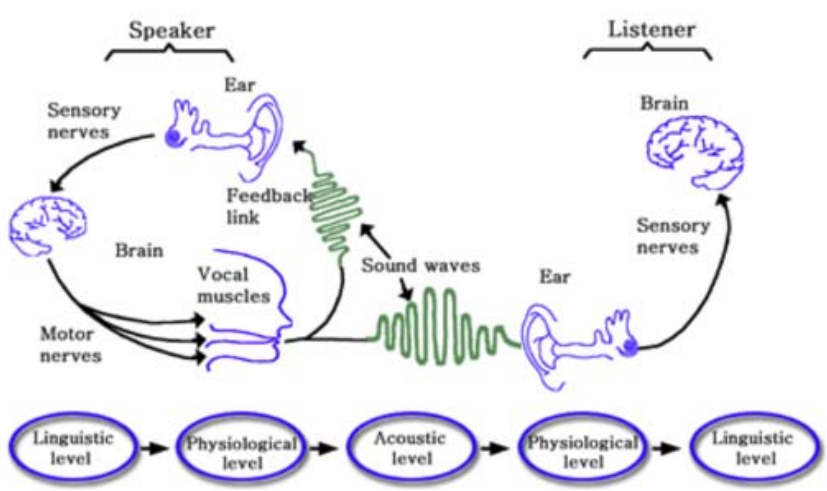


Figure 1. Linguistic information flow in human brain

Actually, human speech sounds are not special more than animal vocalizations and environmental sounds. What is really special emerges when sound waves are *perceived* in human auditory brain system. The categorical perception (Liberman et al., 1957) is a typical example of indicating how speech perception is distinguished from other sound perception. Thanks to animal studies, we already have knowledge enough to account for lots of physiological mechanisms of transforming acoustic waves into neural signals via auditory pathways from cochlea to auditory cortex (see Figure 2).

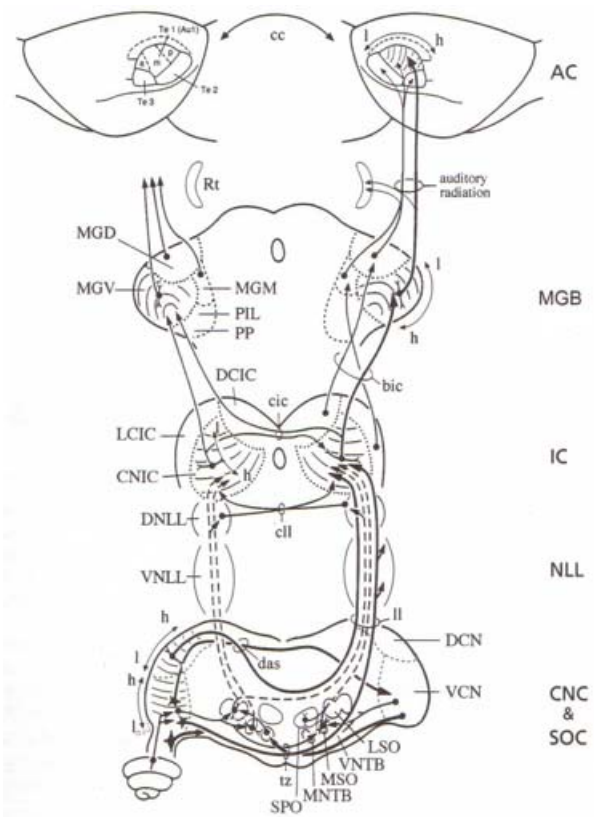


Figure 2. Central auditory pathways in the rat

The sound incoming through outer ear travels along the (ascending) auditory pathways – cochlea nucleus complex (CNC), superior olivery complex (SOC), nuclei of the lateral lemniscus (NLL), inferior colliculus (IC), and medial geniculate body (MGB) – until it reaches auditory cortex (AC) (Malmierca and Hackett, 2010). Each region drives or modulates the others to parse and encode the incoming acoustic waves. At early stage of these pathways, temporal codes synchronized with the incoming sound waves are generally used to faithfully represent the speech sounds as physiological signals, i.e. neural spikes (Gerstner et al., 1997). However, it turns into rate codes at later stage by sensory coding mechanism of hierarchical representations at neuronal level, which are optimized for several sound types (Smith and Lewicki, 2006). Furthermore, specific and nonspecific cortico-thalamo-cortical interactions to facilitate (or inhibit) the incoming sounds are added to shape the auditory codes in detail according to top-down and bottom-up modulating mechanisms (Bachmann, 2006).

In cortical regions, more complex is sound processing modulated not only by the physical properties of sounds, e.g. frequency, amplitude, category, but also by many high-level linguistic factors. According to such findings, for example, we have several models (or linguistic pathways) for neural systems of sound identification and localization as well as specific sound representation in primary and secondary auditory cortices. Among them, voice-selective cortical areas (Belin et al., 2000) are notable in that it is an evidence of species-specific neural circuitry in human, which was also found in macaque monkeys (Petkov et al., 2008). That is,

it seems that there is a point of turning sound waves into some meaningful speeches in the brain.

Now, our questions will be these: what is special for speech in human? What makes it different from other sounds or even other animal vocalizations? How can we learn any sounds with specific meaning? Specifically, how can a sound be associated with a specific concept? And, what neural systems are needed to do that? All these questions are worthy asking in themselves and it is also required to investigate them in that they are very important to reveal early language acquisition, i.e. neural mechanisms describing the whole processes from sound imitation to word learning.

2. From sounds to speech

After the birth, during about one year, an infant experiences several learning phases before her first word production (Kuhl, 2004). The learning is simultaneously initiated in both perception and production in parallel (see Figure 3), basically relying on several sensorimotor integration processes (Hickok, Houde, and Rong, 2011). According to Kuhl (2004), for the whole period in infants, sensory learning is prerequisite to speech perception ability. As the infants are more exposed to native language environment, she comes to have preference of producing speech sounds, e.g. vowel-like sounds, and also have language-specific perception for vowels, accompanying detection of typical linguistic features in her

native language. This is the end of universal speech perception and as a result, there is a decline of discriminating sounds in foreign language.

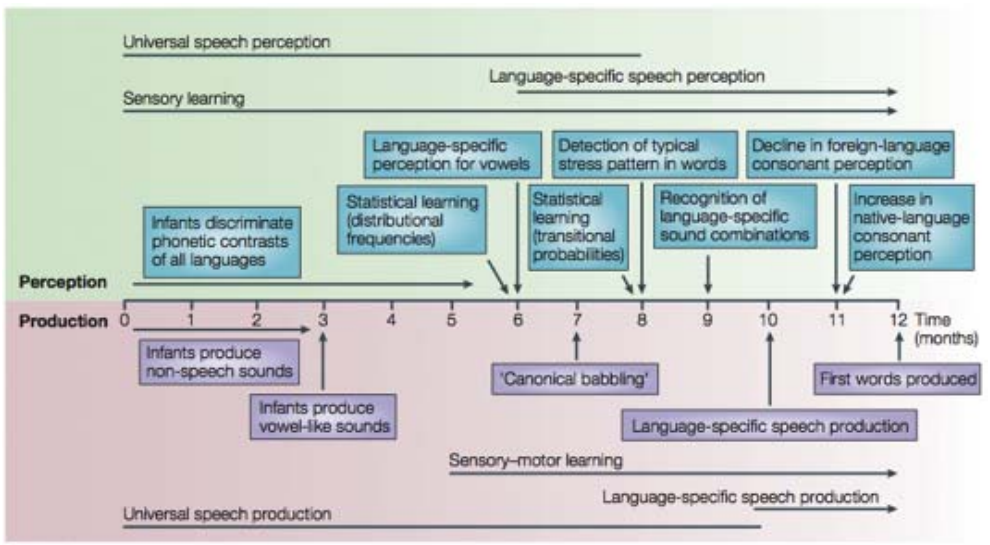


Figure 3. The universal language time-line

Specifically, what happens in the language-specific sound learning? As an example, human auditory system has natural sensitivities concerning the possible voice-onset-time (VOT) boundary in the region of +20 to +40 msec (in English). By the way, chinchilla (a small rodent) also shows a remarkable similarity to native English listeners in its categorization of VOT continua between *ta* and *da* differing in VOT by 10 msec (Kuhl and Miller, 1978). Similar to humans, it supports the existence of a natural boundary of VOT in chinchilla, without rigorous sensorimotor integration for speech perception. That is, the linguistic information embedded in the acoustic sounds is not complete enough to account for speech perception.

The VOT boundaries for both humans and chinchilla are the result of sensory learning, which was associated with meaningful objects or events in their living environments. Consequently, acoustic features should be associated with higher-level linguistic information, i.e. meaning, for successful speech perception. In this vein, language acquisition in infants is a process to continuously associate meanings with specific sounds. It is the key feature to turn sounds into speech, too. For this reason, we aimed to examine the difference between meaningful sounds and meaningless ones during speech processing. We modeled a simplified speech in this study as verbal repetition described in the next section.

3. Speech and verbal repetition

Verbal repetition is basically a sort of vocal imitation and it is divided into several different classes. Basically, imitation simply means sound mimicking without articulating sounds phonetically and phonologically. Importantly, no semantic information is accompanied with the imitation. Similarly, echolalia, repetition of vocalizations, also has no semantic information, but it is an articulated sound by the imitator. Besides, there are several different repetitions (see Table 1). For example, shadowing is an immediate repetition not waiting for the end of words, usually 150 to 250 msec delayed from the speech onset time. Impersonation is widely known as an imitation of the voice or manner of other's speech. Similar to the impersonation, dialect repetition imitates the prosodic

elements of speech as well as essential linguistic information. More specifically, if one uses different pitches with one's own voice, it is falsetto. Subvocal repetition or inner speech has no explicit articulatory movements, while silent repetition has explicit articulatory movements but there is no sound. In case of foreign language repetition, there are two different cases: (unknown) foreign language repetition not evoking phonological or semantic information, and (unfamiliar) foreign word repetition with phonological components, but with no semantic ones.

Table 1. Classes of verbal repetition

Classes	Perception						Production		
	acoustic	phonetic	phonological	morpho-syntactic	semantic	prosodic	articulation	acoustic	prosodic
Imitation	√					√			
Echolalia	√	√	√	√			√		
Shadowing	√	√	√	√			√		√
Impersonation	√	√	√	√	√		√	√	√
Dialect repetition	√	√	√	√	√		√	√	√
Subvocal repetition	√	√	√	√	√				
Falsetto	√	√	√	√	√		√	√	
Unknown Foreign language repetition	√	√					√	√	√
Unfamiliar foreign word repetition	√	√	√				√	√	√
Silent repetition	√	√	√	√	√		√		

In this study, we aimed at examining brain activities during online speech processing, which could reveal language *in situ* exactly. To this end, we introduced (immediate) verbal repetition tasks as a typical example of online speech, because (1) listening and repeating appear to have a significant interaction during early language acquisition (Corrigan, 1980; Iverson et al., 2003; Kuhl et al., 1992); (2)

many neuropsychological data suggest that repetition ability is pivotal in specifying aphasic types – for example, verbal repetition is impaired in conduction aphasias but relatively intact in transcortical aphasias that spare the perisylvian language network (Wallesch and Kertesz, 1993); (3) verbal repetition is simple and fully linguistic, in contrast to other studies requiring nonlinguistic processes. In order to study speech and language, most studies adopted unnatural tasks involving additional mental processes such as matching, detection, discrimination, and judgment (Binder et al., 2000; Cummings et al., 2006; Husain et al., 2006; Poldrack et al., 1999); and (4) we can observe a simple form of *in vivo* language by repetition, since both input and output ends are involved in the repetition. In terms of information processing, the time course of online processing is central to a complete description of behavior (Massaro and Cowan, 1993).

4. Purpose and organization of this study

While investigating neural mechanism of verbal repetition, the purpose of this study is to show (1) how speech sounds are represented in the brain during online speech processing, especially in contradistinction to meaningless sounds; (2) temporal characteristics of neural activities responding to speech sounds compared to other sounds, e.g. natural environmental sounds and animal vocalizations; (3) what neural mechanism makes meaningless speech sounds turn

into meaningful sounds; and (4) a plausible neurolinguistic model of verbal repetition that can account for our results.

To these ends, we conducted three consecutive experiments based on verbal repetition: two functional magnetic resonance imaging (fMRI) experiments and one functional near infrared spectroscopy (fNIRS) one. The organization of this study is as follows: (1) we introduce theoretical background of this study; (2) we demonstrate how speech codes are generated and maintained in the brain; (3) we show hemodynamics of speech processing; (4) we investigate which brain network is recruited for learning speech sounds; (5) we discuss the findings with respect to neurolinguistic model of verbal repetition; and (6) we conclude it with some future remarks.

Chapter 2. Speech representation

In the first experiment, we investigated how speech is represented in human brain during verbal repetition. It is widely known that in terms of neural processing, speech is distinguished from other sounds such as meaningless sounds produced by humans (pseudowords or nonwords), animal vocalizations, natural (or environmental) sounds, and so on. However, most studies failed to rule out the confounding factors at lower sensory level intrinsically introduced by the physical (or acoustic) difference of those sounds. Furthermore, no complete description of online speech has been provided because most studies analyzed only one facet of speech processing by dissociating perception from production, and vice versa. Here, we aimed at examining neural activities when a sound is perceived as a meaningful speech even though the acoustic feature of the sound waves is not changed.

1. How are sounds represented during verbal repetition?

The discovery of Broca's and Wernicke's areas provided substantial grounds upon which to localize linguistic functions within the brain; since then, speech has been studied in two separate and complementarily opposed contexts, those of production and perception (Lichtheim, 1884). It is important to

remember that speech is by nature the result of rigorous sensorimotor integration. In other words, the motor system may be shaped by feedback control of perceived speech and simultaneously may modulate the auditory system during speech perception (Ben-Shalom and Poeppel, 2008; Levelt, 1989; Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967). This feature is more prominent in pre-lingual children who have just started to learn a language (Kuhl, 2004). However, little is known about the neural activities and computational bases of *in vivo* language, as it is difficult to design online experiments with relevant stimuli and responses in a laboratory environment.

We monitored blood oxygenation level-dependent (BOLD) signals from functional magnetic resonance imaging (fMRI) while subjects immediately repeated what they heard. We designed novel auditory stimuli that sounded like ambiguous mixtures of phonologically similar words and pseudowords. This enabled us to contrast the neural activities evoked by acoustically identical but phonologically different sounds. Statistical analyses were performed to reveal both common and distinct neural activities between the two conditions. At the same time, we examined the neural activities associated with speech perception/production and judgment/button-response by separately modeling the hemodynamic-response functions (HRFs) of these two events. With these techniques, our aims were to localize the audition-articulation interface linking speech perception and production and in particular to ascertain the characteristics of the speech codes underlying the verbal repetition of pseudowords versus words.

2. Experimental Design

1. Subjects and Stimuli

Twenty-two native Korean speakers (11 females and 11 males) aged 18-34 years old (mean 22.8 years) participated in this experiment. Written and informed consent was obtained from all subjects before the experiment. All subjects were strongly right-handed as assessed by the Edinburgh handedness inventory (Oldfield, 1971) and had normal auditory ability and no neurological or medical disorders. The experiment was conducted according to protocols approved by the Institutional Review Board of Gachon University of Medicine and Science.

Novel auditory stimuli were designed to compare brain activity during verbal repetition of sounds that were acoustically identical but phonologically different (see Figure 4). The steps in the stimulus generation were as follows: First, 84 word-pseudoword pairs were chosen from two-syllable words and pseudowords, with the criterion that the word and the pseudoword differed by only one vowel, and thus they are phonetically-similar pairs. The position of the different vowel was either at the first or second syllable with the equal probability over all chosen pairs. Second, the selected words and pseudowords were recorded separately into *.wav* format using SoundForge (Sony Creative Software Inc.), as spoken by a male native speaker of Korean. Third, the recording data

were digitally mastered using MATLAB R2008a (The Mathworks, Inc.) with ambient noise and incidental variations removed by subjecting the recordings to the principal component analysis (PCA) and retaining only two or three major components (containing > 55 % of variance) (Jolliffe, 2002). Fourth, the PC's from each word-pseudoword pair, were added and subjected to multiscale PCA (MPCA) simultaneously at different resolution levels (Bakshi, 1998), which had the effects of smoothing and regularizing sounds in the recordings, i.e., removing within-pair and across-pair variations while maintaining fundamental components in the original sounds.

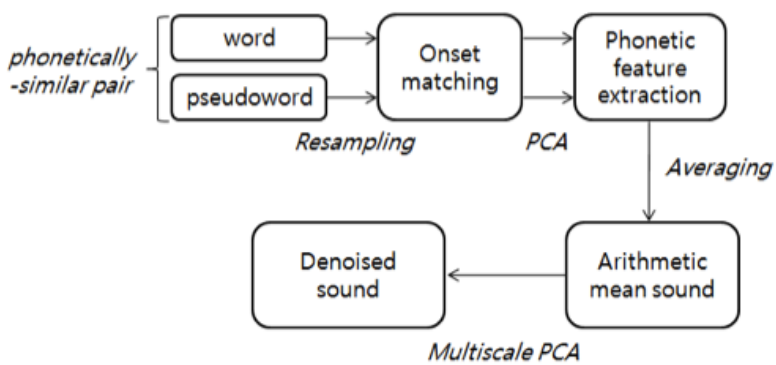


Figure 4. Word-pseudoword mixture generation

Novel stimuli were generated from phonetically similar pairs of a word and a pseudoword. The word-pseudoword mixtures were prepared as computer files in *.wav* format, after several preprocessing to make them ambiguous. A detailed description of the whole process is given in the text.

As the word-pseudoword pairs were selected according to phonological similarity, any semantic ambiguity was not considered here. That is, we controlled semantic distance between words and pseudowords as maximized as possible, which led to minimization of contextual effects in speech perception. When

played back, the reconstructed recordings sounded like a two-syllable speech sound with a vowel ambiguous in the Korean vowel space. Although the ambiguous vowels contained phonetic features common to normal Korean vowels, they were impossible to exactly articulate, since two different tongue positions or mouth openings would have been required at the same time. Perceptually, the situation may be regarded as analogous to what happens with the bi-stable figure-ground segregation in the Rubin vase/profile illusion (Rubin, 2001) or with the bi-stable perceptual interpretations of the Necker cube. One sees the Rubin illusion image visually as a whole, but one cannot perceive both vase and profile images at the same time. Similarly with the Necker cube, one interpretation of the 3-D structure wins out the other interpretation, and never both are perceived simultaneously.

The situation was quite similar with our ambiguous vowel stimuli. When one interpretation of the vowel was perceived, it popped out as the figure and the alternative interpretation faded into the background. Given that acoustic features for the two interpretations were always present in the stimuli and there was no reason to favor one over the other, the perceptual choice would randomly vary from trial to trial. This was in fact observed in subjects' responses (see Figure 5): We evaluated the perception and response of ambiguous stimuli by estimating the subjects' responses. The perceptual judgments were divided equally between the two interpretations, i.e., a word or a pseudoword. The ratio of perceiving a word over a pseudoword was approximately one half in all 22 subjects (the average and standard deviation were 0.52±0.048). The difference between word-

and pseudoword-perceived trials was not statistically significant at the 95 % confidence level (*t*-test, $p = 0.097$; ns: not significant).

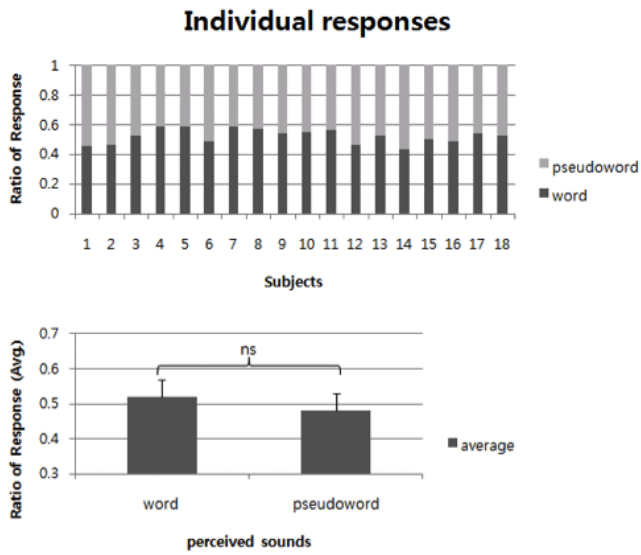


Figure 5. Perception and response of ambiguous stimuli

The perception and response of ambiguous stimuli was evaluated by estimating the subjects' response. The ratio of response between words and pseudowords was approximately close to 0.5 in all subjects (upper graph). The averaged ratio of word-perceived trials was 0.52 ± 0.048 (lower graph). The difference between word- and pseudoword-perceived trials was not statistically significant at the 95 % confidence level (*t*-test, $p = 0.097$; ns: not significant).

In another experiment described in the Appendix, we asked whether the acoustically ambiguous stimuli were processed deeply enough, so that the resulting bi-stable phonological perceptions competed at the lexical level. To this end, ambiguous stimuli were constructed with two words with different frequencies. A lexical effect was observed where the perceptual responses by subjects' were biased favoring higher-frequency words, suggesting that our ambiguous stimuli led to bi-stable phonological perceptions and the competition was influenced by activations at the mental lexicon.

2. Experimental Procedure

The auditory stimuli were presented binaurally inside the MRI system via a magnetically shielded audio system (SS-3100 Silent Scan, Avotec Inc.), and it was the task for subjects first to immediately repeat what they heard while lying in an MRI system and then to report whether they heard a word or a pseudoword by pressing a button. In general, it is hard for subjects to listen and repeat exactly what they heard in a noisy MRI system. Therefore, to prevent the scanner noise from interfering with auditory perception, we employed the interleaved silent steady state (ISSS) sequence, a sparse-imaging method that allows brief silent periods for verbal repetition between image-acquisition pulses (Schwarzbauer et al., 2006). The mixed word-pseudoword stimulus was delivered during a silent pause lasting for two seconds, followed by an additional 2 sec for verbal response. The subjects repeated the stimulus verbally and reported their judgment by pressing a button as the image-acquisition pulses resumed. Subjects were instructed to listen carefully throughout the run and to repeat the stimuli at approximately the same speed. They were also asked to articulate as clearly as possible and in a quiet voice, to minimize head movements. The stimuli and subjects' verbal responses were monitored via a microphone built into the MRI system. According to Jezzard and Clare (1999), 2-3 pixels of local distortion in the echo-planar image (EPI) can result from a 5° flexion. In order to minimize artifacts from the distortion, we tried to keep local distortion parameters within 2° flexion,

in which the local distortion was about one pixel or 4 mm. Therefore, data with distortion greater than 2° flexion were excluded from the statistical analysis, and so this study is largely free from the speech artifact problem in spite of the overt speech. In addition, an event-related fMRI design made it possible to examine brain responses separately for auditory stimulation/verbal repetition and judgment report that might be unnaturally required during the repetition tasks (see Appendix).

A trial lasted for 16 sec and consisted of auditory stimulation (perception phase; 2 sec), verbal repetition (production phase; 2 sec), button response (judgment phase; 2 sec), and rest period (10 sec). A run of MR scanning contained an initial rest period of 12 sec followed by 21 consecutive trials in random order (see Figure 6). Each subject had four separate runs, i.e. 84 trials randomly selected out of total 168 stimuli, with randomized presentation to improve design efficiency and SNR (Signal-to-Noise Ratio) of event-related fMRI design (Dale and Bruckner, 1997; Lindquist, 2008; Lindquist and Wager, 2007; Liu, Franck, Wong, and Buxton, 2001). Subjects listened carefully to a stimulus during the perception phase and then repeated it during the production phase, irrespective of whether they recognized it as a word or a pseudoword. During the judgment phase, they reported their judgment in a two-alternative forced choice (2AFC) manner: the left button for a word and the right button for a pseudoword. Before data analysis, all trials were classified into word- and pseudoword-perceived trials according to the button-press reports. The randomness of subjects' responses was also analyzed to confirm that the ambiguity of the stimuli was not biased toward any

trial conditions. We calculated the respective proportions of word- and pseudoword-perceived trials that subjects responded to, as shown in Figure 5.

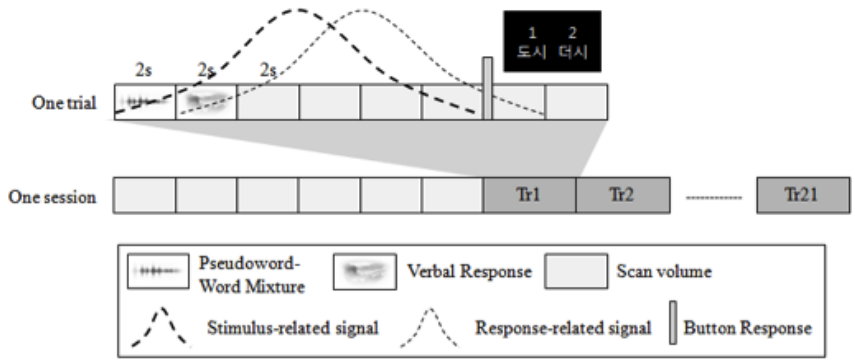


Figure 6. Experiment design for the first experiment

One trial consisted of a stimulus-response pair followed by six scan volumes, in an event-related design. One session consisted of 21 trials with six dummy scans. We asked the subjects to respond to a two-alternative forced choice (2AFC) task by pressing a button after verbally repeating the stimulus.

3. Data acquisition and analysis

A 3T MRI system (Verio, Siemens Medical Solutions) with 12-channel head matrix coil equipped with echo-planar imaging (EPI) capability was used to get MR signals with higher Signal-to-Noise Ratio (SNR) in an event-related fMRI design. T_1-weighted anatomical images were obtained first (TR = 380 msec and TE = 3.06 msec), and then T_2^*-weighted EPI images were obtained at the same slice locations with the following parameters: brain volumes = 132 (n = 132), TR = 2000 msec, TE = 30 msec, flip angle = 90°, field of view = 220 × 220 mm^2, matrix size = 64 × 64 pixels, number of slices = 30, and slice thickness = 5 mm with no gap, parallel to the anterior commissure-posterior commissure (AC-PC) plane. As we

used a sparse imaging method, there were only 132 brain volumes, less than normal measurement for one session (174 volumes for 348 sec). The six images for the first 12 sec were discarded to better approximate a steady state in the MR signal by excluding approach time.

SPM5 (Wellcome Department of Cognitive Neurology) was used for image realignment, Gaussian filtering with 8 mm full width at half maximum (FWHM), and spatial normalization of the brain volumes to the Montreal Neurological Institute (MNI) templates for this experiment. Before spatial normalization, slice timing was corrected to match the different timing of MR signals between the first and last slices. After preprocessing of brain images, we analyzed the individual brain data according to the general linear model (GLM; Frackowiak et al., 1997). Neural activity was modeled by a canonical HRF at trial onset time. The results were mapped on the template MNI brain and rendered on 3-D brain with the help of SPM5 and MRICro (Rorden and Brett, 2000). In addition, fixed effect and random effect with group data were modeled to identify statistically consistent activities in the brain. The threshold value of significance was set at $p < 0.05$ for the whole-brain data and the ROI (region of interest) data. We corrected the threshold with FDR (false-discovery rate; q-value = 0.05) to minimize problems from multiple comparisons. The q-value was selected for each data in advance, to hold the false-discovery rate below 5 % (Genovese, Lazar, and Nichols, 2002).

The ROI areas were defined by referring to our previous experiments and the literature of speech processing (Hickok and Poeppel, 2007; Indefrey and Levelt, 2004; Newman and Twieg, 2001; Vallar et al., 1997). It covered temporal lobe

areas (BA20, 21, 22, 38), primary and secondary auditory association cortices (BA41, 42), Wernicke's area (BA39, 40), prefrontal areas including Broca's area (BA43, 44, 45, 46, 47), cingulate cortex areas (BA23, 24, 31, 32, 33), and premotor and supplementary motor areas (BA6). The MNI-normalized brain mask image including the white matter areas was automatically generated by WFU_pickatlas software (Maldjian et al., 2003), which was based on the Talairach Daemon database (Lancaster et al., 1997; Lancaster et al., 2000). For visualization of activated regions, the MNI coordinates were again converted into the Talairach coordinates by a non-linear transform to get a good match for both the temporal lobes and the top of the brain in all images and rendered on the template brain image (Calder, Lawrence, and Young, 2001; Duncan et al., 2000; Talairach and Tournoux, 1988).

As the repetition task involves a button response or judgment phase that evokes nonlinguistic neural processes, the neural activities by the verbal repetition should be analyzed in isolation from the others. Furthermore, in a rapid event-related design of short ISI (inter-stimulus interval), the design efficiency of fMRI study is significantly reduced as the ISI decreases. For this reason, jittering is generally used to randomize the onset of stimuli (Burock et al., 1998). However, the randomized ISI between two subsequent events, perception and production, are likely to cause all trials not to be equally processed, recruiting unintended cognitive processes in the additional time, e.g. subvocal rehearsals, a short-term memory management, a retrieval of phonologically- and semantically-correlated words. Therefore, we treated perception and production as a single event and

divided the repetition task into two different phases: speech perception/ production and later judgment (button response); we modeled separate HRFs at the onset of each phase in all trials. We made *t* statistical maps of significantly activated voxels for the individual subject brain, in contradistinction to the baseline condition, and then specified brain areas consistently activated across all subjects during the tasks, by applying random effect analyses with group data. While applying a pairwise *t*-test for each voxel (volumetric pixel), the threshold value for significant voxels was relevantly corrected ($p < 0.05$, FDR). From the results, we localized brain areas involved in repetition of each stimulus.

Next, we classified all trials into two groups, namely word- and pseudoword-perceived trials. After localizing the activated areas by trial, we examined pairwise contrasts between the two groups with *t* statistical maps (Friston et al., 1999; Price and Friston, 1997). The result showed which neural areas were commonly activated while subjects repeat all stimuli. In addition, we specified the content-specific activities selectively modulated by word- and pseudoword-perceived stimuli in all activated areas ($p < 0.05$, FDR). The spatial signature of the activated patterns was summarized and further explored to identify which areas were uniquely recruited during the tasks (see Results below). The laterality of repetitions was also estimated by statistically quantifying the asymmetry of activated patterns. The laterality of each stimulus was defined as the ratio of significantly activated voxels between the left and right hemispheres. To this end, we flipped a contrast image at the sagittal plane and overlapped it with a normal image. The results identified which hemisphere was more activated

by repetition tasks. Laterality was evaluated by considering the significance of right-dominance in each stimulus condition (Husain et al., 2006).

3. Results

1. Task-related neural activities

Repetitions of all stimuli commonly activated the perisylvian language network for speech (Catani, Jones, and Ffytche, 2005) and some parts of the frontal and parietal areas as well (see Figure 7).

We found several local maxima in the left and right superior temporal gyri (BA22, 38), left precuneus, left and right medial frontal gyri (BA6), left inferior frontal gyrus (BA9), right superior frontal gyrus (BA6), right middle frontal gyrus, right cingulate gyrus (BA32), left middle temporal gyrus, right lingual gyrus (BA17), left thalamus, and elsewhere (see Table 2). Largely, verbal repetition tasks activated two salient networks, i.e. the fronto-temporal and fronto-parietal networks.

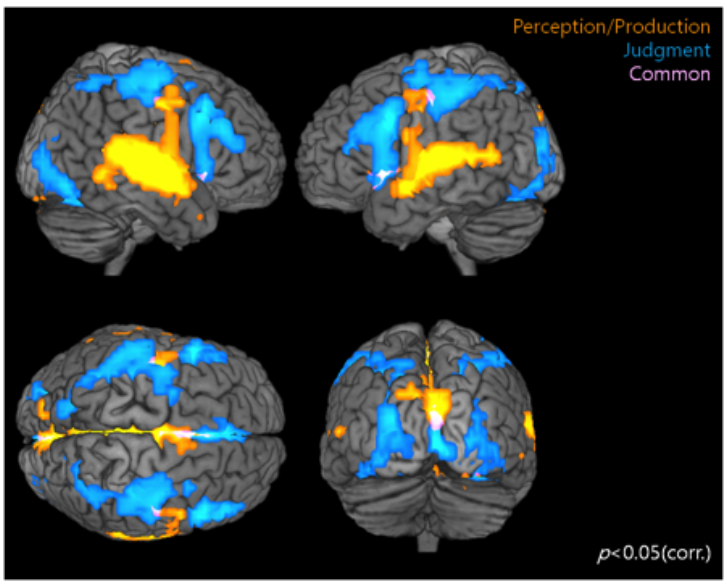


Figure 7. Neural activity accompanied with verbal repetition

Neural activity accompanying repetition was statistically mapped onto the MNI template brain. The repetition task was classified into two phases: perception/production (orange-yellow) and judgment (blue), and the neural activities were separately modeled with hemodynamic-response functions (HRFs) at the event onset of each phase ($p < 0.05$, corrected). The light pink area indicates activity overlap between different phases.

The broad temporal lobe area correlated to auditory processing was bilaterally activated, including the superior temporal and middle temporal gyri, along with the premotor cortex in the frontal area. By verbal repetition, this fronto-temporal network was more activated in the right hemisphere than in the left homologue, implying asymmetric involvement of the right hemisphere during listening and repeating. In particular, the inferior parts of the premotor cortex near the Sylvian fissure were distinctively activated in the perception/production phase.

In the judgment (button-response) phase, three different clusters were found outside the perisylvian region. This fronto-parietal network, in parallel with additional loci in the occipital lobe, showed complex neural activities demanding

visual and motoric processing and some higher-level cognitive functions needed in the judgment phase. This also involved the middle and inferior frontal areas, i.e. the pars opercularis (BA44), anterior to the premotor cortex, localized in the perception/production phase, and the superior parietal lobe area next to the motor cortex bilaterally. The neural activities identified in the occipital cortex apparently corresponded to visual processing in the judgment phase. Unlike during the perception/production phase, however, the neural activities in the right hemisphere were not prominent in the judgment phase.

Table 2. Local maxima of activated brain regions

Brain Regions	**Brain Regions**	***x***	***y***	***z***	**peak *t***
Word-perceived (Perception/Production)	L Precuneus	-8	-73	22	11.06
	L Superior Temporal Gyrus	-63	-27	6	10.79
	R Superior Temporal Gyrus	51	-23	6	10.72
	R Frontal Lobe	36	17	22	6.27
	L Anterior Cingulate	-8	32	21	4.12
	R Superior Temporal Gyrus (BA22)	63	-4	0	6.38
	L Medial Frontal Gyrus (BA6)	-4	3	60	6.29
	L Superior Temporal Gyrus (BA22)	-63	-4	5	5.96
	R Tuber	44	-56	-27	4.99
	R Temporal Lobe	48	-20	-16	4.19
	L Superior Temporal Gyrus (BA38)	-48	19	-18	3.96
Word-perceived (Judgment)	R Lingual Gyrus (BA17)	4	-85	4	10.13
	R Middle Frontal Gyrus	51	21	36	7.65
	L Inferior Frontal Gyrus (BA9)	-55	13	32	8.45
	L Superior Frontal Gyrus	0	10	50	8.15
Pseudoword-perceived (Perception/Production)	R Superior Temporal Gyrus	51	-23	6	6.65
	L Superior Temporal Gyrus	-63	-27	6	6.58
	L Middle Temporal Gyrus	55	-47	-6	5.60
	L Precentral Gyrus	-16	-20	65	4.24
	R Superior Temporal Gyrus	63	4	0	7.29
	L Superior Temporal Gyrus (BA22)	-63	-8	0	5.38
	R Midbrain	16	-16	-12	4.46
	R Cerebellar Tonsil	36	-52	-31	4.21
	R Superior Frontal Gyrus (BA6)	4	7	60	4.12
	R Cigulate Gyrus (BA32)	4	25	26	4.06

Pseudoword-perceived (Judgment)	R Inferior Frontal Gyrus	51	13	27	9.41
	L Postcentral Gyrus (BA3)	-36	-21	47	8.47
	L Medial Frontal Gyrus	0	6	46	8.17
	L Thalamus	-12	-19	6	7.42
	R Middle Frontal Gyrus	51	21	36	6.57
	R Precuneus	8	-48	53	4.41

Only activations with a $p < 0.05$ (corrected) and a volume of at least 640 mm^3 (10 measured voxels) were considered. The x, y, and z values show the center of gravity of the activated clusters in Talairach coordinates. L, left; R, right; BA, Brodmann area.

2. Word- versus pseudoword-perceived neural activities

Next, we divided all trials into two conditions, namely, word- and pseudoword-perceived trials, according to the subjects' perception, and then contrasted their neural activities. From the results, we found distinct neural activities modulated by the subjects' perception not only in the perception/production phase but also in the judgment phase (see Figure 8).

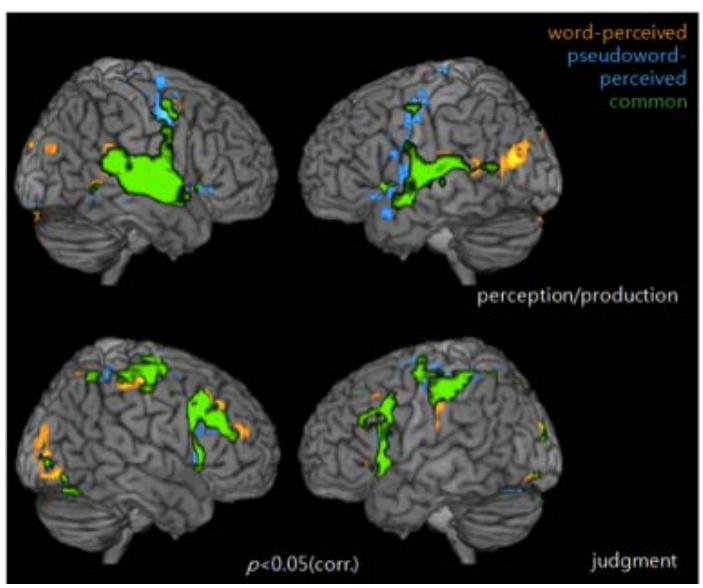


Figure 8. Neural activity by different conditions and phases

The pairwise contrasts between word- and pseudoword-perceived trials show distinct neural processes for word-perceived (orange-yellow) and pseudoword-perceived (blue) trials evoked at perception/production and judgment phases respectively ($p < 0.05$, corrected). Activities common to both conditions are indicated in green.

The orange-yellow region was selectively activated by the word-perceived trials only, whereas the activations at the blue region were specific for the pseudoword-perceived trials. We examined the BOLD signal changes further at the activated clusters to see whether the two types of trials recruited different networks or the same network but to different degrees (Koelsch et al., 2009).

In Figure 9, brain activity during word-perceived trials was plotted for the perception/production and judgment phases in the upper and lower panels, respectively. Note that the BOLD signal increased at the word-specific activation (the curve marked with diamonds) as expected, whereas it decreased at the pseudoword-activated area (the curve marked with squares), indicating that the two regions were distinctly recruited for word- versus pseudoword-perceived trials. Similarly, the opposing pattern of BOLD changes was also observed between the two cortical areas for the pseudoword-perceived trials (data not shown).

These results indicated that the word- and pseudoword-perceived trials activated distinct networks of areas (i.e., orange-yellow and blue regions in Figure 8), rather than the same network at varying degrees. Moreover, activations associated with the two trial-types even differed across hemispheres. As shown in Figure 8, word-perceived trials evoked lateralized activities in all phases: left-lateralized in the perception/production phase and right-lateralized in the

judgment phase. However, no such lateralization was significantly found in pseudoword-perceived trials.

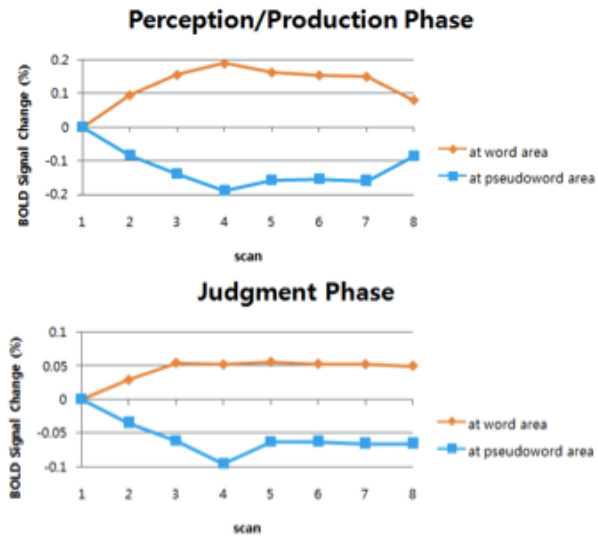


Figure 9. Relative BOLD-signal change for word-perceived trials

The neural activities by word-perceived trials were plotted in both word- and pseudoword-specific regions. The relative BOLD signal change was defined as percent signal change of each trial in contrast to baseline condition. For comparison, the relative BOLD signal change was averaged across all trials and normalized to the value at trial onset.

Notably, we found a fronto-temporal dichotomy between pseudoword- and word-perceived trials. In the perception/production phase, the pseudoword-perceived trials recruited premotor cortices bilaterally while the word-perceived ones recruited the posterior parts of the left superior temporal sulcus and angular gyrus but not the right ones. In the frontal areas, we found no unique neural activities accompanying the word-perceived trials. Instead, there were neural activities accompanying the pseudoword-perceived trials in frontal areas such as the inferior frontal gyrus near BA44 and premotor cortices in the left hemisphere.

The same fronto-temporal dichotomy was not found in the judgment phase. In both trial conditions, the judgment phase recruited the fronto-parietal network bilaterally. The lateral-occipital cortex close to BA18 and 19 was uniquely activated by word-perceived trials in the right hemisphere. We also noted this hemispheric difference of neural activities, determined by different trials and phases, in the repetition task. The quantitative data indicates that the hemispheric difference was greater in word-perceived trials than in pseudoword-perceived ones. In terms of processing phases, neural activity of verbal repetition tasks showed hemispheric asymmetry in the perception/production and judgment phases respectively.

We explored further how neural activities involved in the repetition task changed according to the two phases of the task given above. The results showed a functional shift of neural activities between sensory and motor representations of incoming speech sounds. From perception to production, the neural activities accompanying the processing of the incoming speech sound moved from the temporo-parietal region, part of the ventral stream, to the frontal region. In the judgment phase, it was evident that the visual and motor areas were activated by the task. The activities shown in the frontal and parietal region seem to be correlated with inner speech and high-level cognition to select a response from two stimuli shown on the screen. The perisylvian region was consistently activated through the whole tasks, indicating that it might be a fundamental linkage between perception and production (a common audition-articulation interface).

4. Discussion

1. Verbal repetition of ambiguous speech sounds

Since we introduced novel stimuli designed to remove acoustic difference, any specific neural activity can be assumed to be relatively free from acoustic processes and instead more correlated with perceptual differences in stimuli. The perceptual process of ambiguous speech is not clearly known yet. Therefore, the contrasts between word- and pseudoword-perceived stimuli in this study have a weakness to some extent. For instance, it might not be enough to allow the discrimination of the phonological process from other perceptual processes evoked by lexical and semantic difference. However, to assess this possibility, we have to investigate speech in isolation from the linkage between audition and articulation to control for these differences; this is of course unsuitable for our aim of examining the two ends while they are simultaneously active.

Fortunately, the verbal repetition task is done over a very short time period and is by nature phonological in that the task does not demand higher-level cognitive processes such as recognition or discrimination of incoming sounds. We successfully designed a suitable natural environment by separating the judgment phase from the verbal repetition phase. We placed the judgment phase after the end of the audition-articulation loop and carefully modeled it with distinct HRFs. As a result, little correlation was found between the judgment and

repetition phases, as shown in Figure 7. The lexical or semantic processes implicitly involved in this study are likely in this sense to be a part of building relevant speech codes. Therefore, it is more plausible to conclude that perceptual processes are normally accompanied by phonological process to build speech codes as a whole, but they are differentially applicable to individual sounds.

Another concern about the ambiguous sounds is that unattended auditory words can access both lexical and semantic representations to some extent (Sabri et al., 2008; Underwood, 1981; Yates and Thul, 1979). When two stimuli with the same or different modalities are presented at the same time, the unattended stimulus seems to have a substantial effect on cognitive activity, since two separate cognitive processes are automatically recruited from the early stage of sensory processing. In this study, however, the subjects listen to only one sound at a time, meaning that only a single attentive process is recruited at the auditory level. This kind of bottom-up attention seems not to have a pivotal role in this study, since the acoustic features of all stimuli were identical for word- and pseudoword-perceived conditions. Instead, higher-level perceptual processes are inextricably accompanied by top-down attention that is inferred only later (Bonte et al., 2006; Jacquemot et al., 2003; Noesselt, Shah, and Jäncke, 2003). Therefore, it is hard for the present study to consider possible subliminal effects evoked by the unattended auditory words.

2. Spatiotemporal localization of neural activities and its implications

Overall, neural activity associated with verbal repetition was found equally in both hemispheres but in weakly right-lateralized patterns, probably due to the feedback-control loop for speech perception and production (Bozic et al., 2010; Tourville, Reilly, and Guenther, 2008). However, this changes as verbal repetition proceeds. The temporal change in neural activity was localized in different regions. It was evident that there was a transition from the temporal to frontal area; nevertheless, some neural areas, for example, the middle parts of the superior temporal and precentral gyri and the supplementary motor area were commonly involved in the perception/production phase. This implies that the transition from perception to production is continuous, so that these processes interact rigorously during speech processing. However, what is notable about the transition is that the premotor and supplementary motor areas in the frontal lobe were activated before speech production. This is reminiscent of the listening effects on the excitability of tongue muscles described in Fadiga, Craighero, Buccino, and Rizzolatti (2002), indicating that sound perception involves motor processes (Cohen, Grossberg, and Stork, 1988). In the judgment phase, neural activities moved into other areas, as the judgment task required very different neural processes: visual reading, including inner speech, and motor response, in order to select one button.

Interestingly, word- and pseudoword-perceived trials were different in terms of the temporal profile of unique activation. The word-perceived trials showed unique neural activities in the posterior parts of the left middle temporal gyrus during the perception/production phase, whereas the pseudoword-

perceived trials showed unique neural activities in the left inferior and superior frontal gyri during the same phase. The word-specific activities partially overlap with the human voice-selective brain regions suggested by Belin et al. (2000). They are also a part of the phonological and combinatorial networks (Hickok and Poeppel, 2007) and indicate that semantic involvement has a certain role in perceiving the human voice, in addition to the phonological characteristics of sounds. No such semantic involvement was found in the pseudoword-specific activities. Instead, a considerable amount of neural activity was found in the precentral gyrus, indicating that motoric involvement becomes important in perceiving meaningless human vocalizations such as pseudowords. Most neural activity found in the frontal area seems to be related to speech production. There was little word-specific activity, while pseudoword-specific activities were easily found.

With respect to the finding of greater brain activation from pseudowords than from words, one possible explanation is that the brain regions involved in lexical recognition and semantics processed real words more easily (Herbster, Mintun, Nebes, and Becker, 1997; Price et al., 1994; Rissman, Eliassen, and Blumstein, 2003). This may raise a question about the word-specific activities observed in this study. If the same brain regions are involved in processing both words and pseudowords, there should be similar neural activities in varying degrees in these regions, irrespective of sound content. However, the relative BOLD signal change in these regions showed completely opposite patterns for words and pseudowords (see Figure 9). This means that the word-specific areas

are dissociated from the pseudoword-specific ones in terms of regional cerebral blood flow (rCBF), indicating that separate neural circuits exist for words and pseudowords. Furthermore, this seems not to be the result of the neural efficiency assumed in the semantic processing of words, because verbal repetition requires little semantic processing, for either words or pseudowords.

It is interesting to note that the neural activities found in the left inferior frontal gyrus were similar to those found in Poldrack et al. (1999), who reported on the phonological processing of pseudowords. Specifically, in both studies, verbal repetition of words and pseudowords is understood respectively as semantic recitation (repetition of semantics and phonology) and repetition of phonology. However, semantic involvement in phonological processing seems not to be correlated with semantic knowledge of representations per se. Instead, it is likely to serve as part of a semantic working memory system modulated by lexicality (Gabrieli, Desmond, Demb, and Wagner, 1996). In the framework of verbal working memory, therefore, the same phonological loop is reserved for words and pseudowords as a part of the short-term memory system. Importantly, the loop is divided into multiple cooperative subsystems that are automatically activated by the semantic content, acoustic content, or both of auditory sounds. Accordingly, there is more than one class of speech codes to represent various sounds, as will be discussed below. This diversity of speech codes relies in part on the fact that online speech processing is modulated by top-down and bottom-up information in parallel (Bonte et al., 2006; Jacquemot et al., 2003; Noesselt et al., 2003).

When it comes to functional asymmetry, it has been reported that spoken language is bilaterally processed, in contrast to written language; in other words, spoken word comprehension shows bilateral activity (Bozic et al., 2010). Similarly, we found little hemispheric difference between the tasks. However, the brain asymmetry is widely observed at the anatomical, physiological, and behavioral levels (Denes and Pizzamiglio, 1999; Zaidel, Clarke, and Suyenobu, 1990). In this context, we can understand hemispheric asymmetry in terms of unique neural activities accompanying word-perceived and pseudoword-perceived stimuli. In the perception/production phase, neural activities were left-lateralized for both stimuli: words in the temporal lobe and pseudowords in the frontal lobe. That is, speech perception and production are likely to rely on various language functions that are lateralized on specialized anatomo-functional structures (Soroker et al., 2005). The left-dominance in the inferior frontal gyrus around the perisylvian region was more prominent in pseudoword-perceived sounds than in word-perceived ones. This finding fleshes out the hemispheric dual-route model for lexical and nonlexical processing (Weekes et al., 1999).

3. Multiple speech codes for vocabulary learning by imitation

The repetition task requires the activation of a phonological loop to repeat incoming sounds properly. In Baddeley's sense, the phonological loop is an articulation-based device with limited capacity (Baddeley and Hitch, 1974). It thus operates with individual phonemes and syllables at the phonological level. As

observed in this study, however, verbal repetition is not simply speech-based but language-based, in that it involves semantic retrieval as well as phonological processing. Furthermore, the same speech sounds are perceived as differentially produced by higher-level perception, indicating that speech processing in the phonological loop varies according to the perceptual information available. Therefore, it is possible to suppose that the phonological loop has a substantial role in language learning as a general device to process auditory sounds in any category (Baddeley, 1998; Baddeley et al., 1998; Papagno, Valentine, and Baddeley, 1991). The sensory-level modality effect found in the phonological loop supports the existence of this uncategorized auditory processing maintained in the same loop (Crowder and Morton, 1969).

Sounds in different categories may be represented as various speech codes to be processed in the phonological loop. In terms of language learning, these multiple speech codes are likely to be resolved into a hierarchical speech representation, from simple sound to complex verbal language. We learn to speak primarily by imitating others' utterances. Human speech is built on such imitated and further articulated sounds, each of which is mapped onto a specific meaning afterwards. Early on, before any specific meaning is associated with sounds, a baby first learns to use its vocal organs by simple vocal play, followed by babbling. The outcome of this articulatory learning is a phonological code that encodes the lexeme of a prospective word. The lexeme is important as an entry into the mental lexicon, associated with the lemma (Caramazza, 1997; Levelt, 1992; McClelland, Mirman, and Holt, 2006). Pseudowords do not have such

phonological entries in the mental lexicon, since they are not registered as words. Thus, they have a tendency to be more dependent on the articulatory code generated by imitating sounds. However, word learning can change this situation by repeatedly carrying out a dynamic procedure to develop a speech code for a particular word. This learning procedure is multifaceted, meaning that speech codes are specified by perceiving the phonological features of sound sequences as well as relevant meanings, but at the same time, speech codes are regarded as abstract gestures intended to articulate corresponding sounds (Liberman and Mattingly, 1985; Liberman and Whalen, 2000). In other words, the phonological loop should develop speech codes specified at multiple linguistic levels during word learning, by activating a feedback pathway from perception to production and vice versa.

To account for the interactive process, therefore, it is plausible to suppose heterogeneous speech codes with appropriate speech representations in the phonological loop. As a model of word learning, the proposed system features a common audition-articulation interface and cooperative phonological subsystems specialized for various sounds. The common interface formulates a feedback loop to process uncategorized sounds. Verbal repetition in the loop incrementally involves the cooperative phonological subsystems to build relevant speech codes for corresponding sounds. In the early stages of word learning, sound imitation in the audition-articulation loop is important to build exact phonological codes. Once learned, the phonological codes are further associated with specific (semantic) concepts. The neural activities observed in this study imply a neural change in the

development of speech codes: acoustic-phonetic speech codes for word-perceived trials in the posterior parts of the temporal area (Hickok and Poeppel, 2004) and articulation-based speech codes for pseudoword-perceived trials in the left inferior frontal gyrus.

Notably, the loci of sound imitation are localized within the inferior frontal and superior temporal gyri, which are major parts of the mirror neuron system used in imitative learning (Rizzolatti and Arbib, 1998), implying that the neural circuits used for articulation-based speech codes might develop out of the sound imitative system. Moreover, the common audition-articulation interface identified in this study is very similar to the neural circuits of imitation as seen in Iacoboni (2005) and Iacoboni and Dapretto (2006). All these findings are consistent with the notion of an analysis-by-synthesis facility provided in the speech motor loop, in which the speech coder estimates coding parameters from the original speech signal (Cohen et al., 1988). In this way, the verbalization of imitative sounds after learning is likely to enhance neural efficiency. This sheds light on the role of this system in vocabulary learning by imitation (Gallese and Lakoff, 2005).

In summary, we suggest the existence of a cooperative system with a common audition-articulation interface and a few phonological subsystems. In the system, speech codes are developed from simple imitative codes to detailed, hierarchized phonological and semantic codes largely dependent upon the imitative learning system used for sound learning. This system is a language-based processor, in that incoming sounds are automatically processed in tandem with the concepts available in the mental lexicon. Simultaneously, it is a speech-based

processor, as it processes various sounds in the phonological loop (Burgess and Hitch, 1999).

Chapter 3. Speech hemodynamics

We have investigated how speech sounds are differentially represented in the brain and which brain regions support such transforming process from sounds to speech. In the second experiment, we further analyzed hemodynamics of speech processing in human brain, mainly within inferior frontal gyrus (BA47) that was identified as articulatory speech codes in the first experiment. Thanks to functional near-infrared spectroscopy (fNIRS), we could monitor hemoglobin (Hb) concentration changes in the loci, which provided us with how speech could be distinguished from other sounds such as natural sounds, animal vocalizations, and human emotional sounds as well as how meaningful speech (words) and meaningless speech (pseudowords) could be distinguished from each other in terms of speech hemodynamics.

1. What is the hemodynamic difference between speech and nonspeech?

In the previous experiment, we found that speech codes in the brain are differentially maintained from the others when the incoming acoustic waves are perceived as meaningful sounds. Speech perception is a kind of sound perception, in which linguistic information such as phonetic features, voice-onset-time, its

associated concepts, and grammar are extracted from the sound through sound parsing.

There are several theories of speech perception largely divided into two distinct theoretical perspectives. In the first stance, they basically assume an acoustic representation of speech sounds in the brain, defined as number of acoustic characteristics (Stevens and Blumstein, 1981; Massaro, 1987; Goldinger, 1997; Johnson, 1997; Coleman, 1998). In the second one, speech perception is not a problem in an acoustic domain, but in an articulatory domain. They argue that speaking and listening are both regulated by the same structural constraints and grammar and listener can perceive articulatory movements to generate actual sounds (Liberman and Mattingly, 1985; Fowler, 1986).

The divergence of speech codes, i.e. articulatory and acoustic-phonetic codes observed in the first experiment partially support the notion of speech perception in an articulatory domain. Specifically, articulatory speech codes were activated for unlearned sounds (pseudowords), whereas acoustic-phonetic codes were used for learned sounds (words). It is also intriguing in that the locus of articulatory codes (left inferior frontal gyrus) is a part of the mirror neuron system (MNS) known as a core network of movement imitation (Iacoboni, 2005; Iacoboni and Dapretto, 2006).

In the second experiment, we aimed to investigate whether the locus of articulatory codes are modulated while the subjects perceive sounds in different categories. To this end, we observed hemodynamics at left and right inferior frontal gyri (BA47) using fNIRS during passive listening of various auditory stimuli.

In addition, we observed hemodynamics of the selected loci during verbal repetition of words and pseudowords, which could reveal speech hemodynamics of meaningful sounds in contradistinction to meaningless ones. It is also a natural situation similar to word learning, where infants mimic human speech sounds out of various environmental sounds

2. Experimental Design

1. Subjects and Stimuli

Fifteen native Korean adults (9 males and 6 females) aged 19-37 years old (mean 25.3 years) participated voluntarily in the present study. Informed consent was obtained from all participants before the experiment. All participants had normal auditory ability and reported no neurological deficits. The subjects completed a questionnaire that assessed their handedness, according to the Edinburgh Handedness Inventory (Oldfield, 1971), and all were strongly right-handed (scored 80 or higher).

The auditory stimuli in five different categories classified by their linguistic structures were prepared for this study: natural sounds, animal vocalizations, human emotional sounds, pseudowords, and words (see Table 3). The natural sounds were selected from the Pittsburgh Natural Sounds dataset recorded by Laboratory for Computational Perception and Statistical Learning (CNBC Lab.,

Carnegie Mellon University, USA). It consisted of ambient sounds (rain, wind, streams) with acoustic transients (snapping twigs, breaking wood, rock impacts) around the Pittsburgh region. Recording was carried out using a M-Audio's MobilePre-USB 16-bit/48 KS/s USB-powered Microphone Pre-amp, with all recordings made at 44,100 Hz. Twenty sound files out of the dataset were selected and then cut to be two-second-length with normalized loudness as *.wav* files.

Table 3. Categories of auditory stimuli

Category	Linguistic meaning	Linguistic segment	Same species	Vocally produced	Sound
Natural sounds	-	-	-	-	-
Animal vocalizations	-	-	-	+	+
Human emotional sounds	-	-	+	+	+
Pseudowords	-	+	+	+	+
Words	+	+	+	+	+

The animal vocalizations were collected from Avisoft Bioacoustics, Germany. It covered various animal vocalizations such as monkey, bird, sheep, horse, frog, etc. The recordings were made using SENNHEISER microphones K3/ME80, ME88, K6/ME62, 64, 66 or MKH60 connected to either a SONY DAT recorder TCD-D3, Marantz PMD 671, TASCAM DR-1, HD-P2, SONY PCM-M10, PCM-D50 or Fostex FR2-LE. We again selected twenty sound files from the set: monkey (4 ea), sheep (1 ea), horse (1 ea), dog (4 ea), wolf (1 ea), mice (2 ea), birds (3 ea), frog (2 ea), and bat (2 ea). All files were cut to be two-second-length and normalized as *.wav* files.

The human emotional sounds were collected from the web. We used twenty sound files, consisting of gasp (2 ea), giggle (2 ea), slurp (2 ea), burp (1 ea), cry (1 ea), yawn (2 ea), kiss (2 ea), slurp (2 ea), snore (2 ea), breathe (1 ea), scream (1 ea), and cough (2 ea). All were recorded as *.wav* files and normalized with the same length (2 seconds).

The pseudowords were generated by randomly combining several consonants and a vowel (/*a*/) in Korean, and thus have no meaning in Yonsei Korean Corpus 1-9 (Yonsei Korean Dictionary, 1998). The words were selected from the same Corpus, with balanced word frequency. All pseudowords and words were four syllable lengths. The pseudowords and words spoken by a female Korean native speaker were recorded and converted into computer files of *.wav* format (22,050 Hz, 16bit, stereo). The loudness (average RMS level) of all stimuli was normalized (-60 to 0 dB) by a sound software (SoundForge; Sony Creative Software Inc.).

All stimuli were not significantly different in loudness and did not exceed two seconds in total length. As shown in Table 3, the stimuli were classified in terms of several linguistic features, i.e. whether they have linguistic meaning, whether there is linguistic segment, whether they are produced by same species (aka human), whether they are vocally produced, and whether they are acoustic sounds.

2. Experimental Procedure

Lying in a table, the subjects were asked to repeat what they heard binaurally via an ear microphone in case of pseudowords and words, and otherwise simply listen to the stimuli. The sound volume was relevantly adjusted for comfortable and clear listening. In one category, twenty stimuli were used and totally 100 different stimuli in five different categories were presented to the subjects. The auditory stimuli in five different categories were pooled and then randomly presented to the subjects in four runs (twenty-five stimuli for one run). One trial consisted of two seconds of perception, two seconds of production (only for pseudowords and words), and twelve seconds of resting to avoid interference from other trials (see Figure 10). Therefore, the length of one session was 416 seconds, including initial dummy 16 seconds (totally 6 min 56 sec). There was no production phase for natural sounds, animal vocalizations, and human emotional sounds.

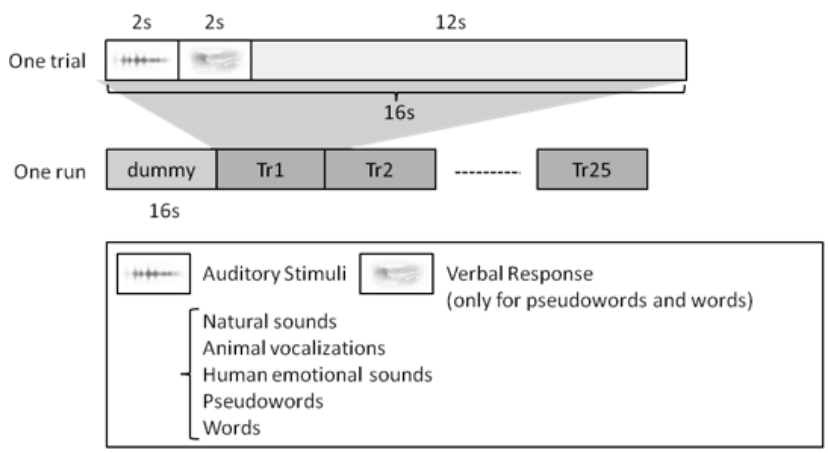


Figure 10. Experiment design for the second experiment

Similar to the first experiment, the subjects were asked to repeat what they heard for words and pseudowords. For the other stimuli, i.e. natural sounds, animal vocalizations, and human emotional

sounds, they simply listened to the stimuli. One trial was 16 seconds in length and one run consisted of twenty-five trials. There were four separate runs.

3. Data acquisition and analysis

During the tasks, the hemodynamic change of left inferior frontal gyrus (BA47), which was identified as the locus of articulatory code recruited during verbal repetition in the first experiment, and its right homologue as experimental control, were monitored by near-infrared spectroscopy (NIRS). We used Oxymon Mark III 8-channel system with sampling rate of 250 Hz (Artinis, The Netherlands), which is capable of measuring the oxygenated (O_2Hb) and deoxygenated (HHb) hemoglobin concentration changes of the optical paths in the brain between the nearest pairs of transmitter and receiver (see Figure 11).

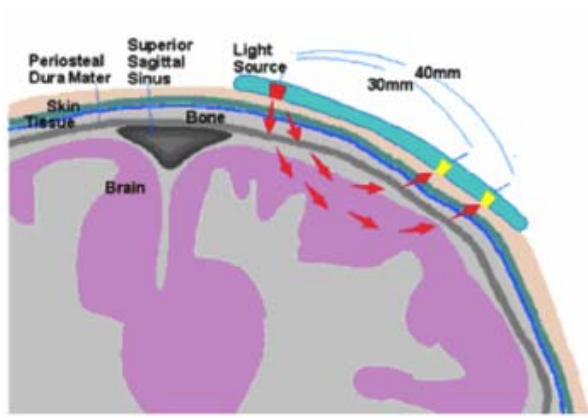


Figure 11. Signal detection between receiver-transmitter in NIRS

The travelling pathways of light are determined by the distance between transmitter and receiver, source wavelengths, characteristics of medium (tissue), and so on. The detecting depth was relevantly corrected to focus on the deep gray matter in inferior frontal gyrus in this study.

The NIRS emits 2 wavelengths (763 and 860 nm) of continuous near-infrared lasers. We introduced 4x1 configuration (see Figure 12), each of which was modulated at different frequencies to detect O_2Hb and HHb at two different brain areas, i.e. left and right inferior frontal gyri (BA47). The activated locus in the experiment 1 (LIFG, [-22 18 -22]) was translated into a coordinate of the 10-20 system (10/20 [-1.9 0.87]) on the scalp surface by Münster T2T-Converter (NRW Research Group for Hemispheric Specialization, Münster University).

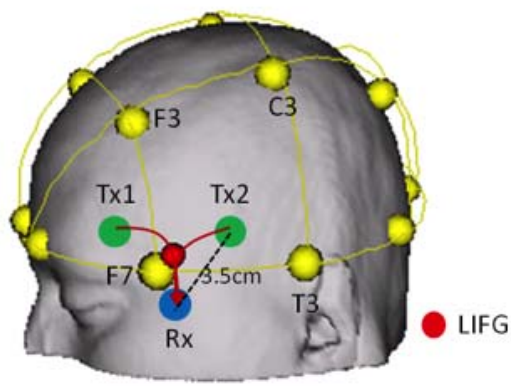


Figure 12. Locus for NIRS monitoring (only left side was shown.)

To detect Hb concentration changes of LIFG, we positioned one receiver and two transmitter optodes near inferior frontal gyri (BA47) bilaterally. The transmitter and receiver were separated by 3.5cm from each other.

To detect the hemoglobin concentration changes of the loci, we separated the distance between transmitter and receiver by 3.5 cm on the scalp surface (see Figure 12), and used differential path length factor (DPF) of 4, by which we could measure hemodynamic changes in the gray matter on the inner brain (Fukui et al., 2003). In the modified Beer-Lambert law (Cope and Delpy, 1988), the

absolute tissue blood volume in ml/100g (TBV) and absolute blood flow in ml·100g^{-1}·min^{-1} (BF) were given as:

$$TBV = \frac{\Delta(O_2Hb - HHb)}{2 \cdot R \cdot \Delta SaO_2} \cdot c_{Hb} \cdot \rho_\tau \cdot k$$

where c_{Hb} (mM) is the hemoglobin concentration of whole blood, $\rho\tau$ (g/cm^3) the specific density of the tissue, *k* a constant reflecting metric conversions, SaO_2 arterial saturation, and R is, in case of cerebral tissue, the large-to-small vessel hematocrit ratio with a value of 0.69 (Lammertsma et al., 1984).

$$BF = \frac{K \cdot \Delta(O_2Hb)}{cHb \cdot 10^{-2} \cdot \int_0^t \Delta(SaO_2)dt}$$

where *K* is a constant representing the molecular weight of hemoglobin, the tissue density and a metric conversion factor (Edwards et al., 1988).

The acquired data were analyzed by the followings: We first extracted time-series data at each optode site, in which noises and motion artifacts were removed by low-pass filtering at 10 Hz (5th-order Butterworth filter). The concentrations of oxygenated (O_2Hb) and deoxygenated (HHb) hemoglobin were then calculated from the above equations. The O_2Hb and HHb data were aligned at stimulus-onset-time to obtain event-locked trials data. With these data, we collapsed all 5 conditions and discovered the time course of the hemodynamic-response function (HRF) to auditory stimuli at each optode site. Compared to the canonical HRF, we described the characteristics of the obtained HRFs. The canonical HRF was based on finite impulse response (FIR) functions (Friston et al., 1995).

Next, we analyzed the HRFs to each of the five conditions at all optode sites and specified the hemodynamic difference between five conditions. We compared speech sounds (words and pseudowords) with nonspeech ones (natural sounds, animal vocalizations, and human emotional sounds). The signal difference was confirmed by statistical analysis in intervals from onset to peak point. We further investigated the change of O_2Hb compared to total Hb (O_2Hb and HHb) at each optode site, which revealed the difference of cerebral blood flow at each locus.

Last, we investigated temporal characteristics of NIRS signals. Though the event-locked NIRS signals showed a single HRF, similar to fMRI signals, the NIRS signals actually have small fluctuations modulated by blood pressure (BP) change, consisting of systolic and diastolic pulsation, respectively. Thanks to high temporal resolution of fNIRS, we could discriminate systolic phase from diastolic one and determine how signal change of each phase contributes to the HRF. To this end, we extracted individual peaks from the NIRS signals and calculated systolic and diastolic NIRS signals by sampling at every positive and negative peak (Lerch et al., 2012). The intermediate samples are interpolated by cubic spline interpolation to obtain the same sample rate of the original NIRS signals. Then, we examined Hb change with both systolic and diastolic NIRS signals.

3. Results

1. Hemodynamic responses at inferior frontal gyri (BA47)

We first measured the hemodynamic responses of speech (pseudowords and words) and nonspeech sounds (natural sounds, animal vocalizations, and human emotional sounds) at inferior frontal gyri (IFG, BA47) bilaterally during the tasks. The results were depicted in Figure 13, together with the canonical HRF (Friston et al., 1995). Similar to the canonical HRF, the normalized O_2Hb changes were significantly increased at the LIFG for both speech and nonspeech, whereas there were significant decreases of O_2Hb change at the RIFG.

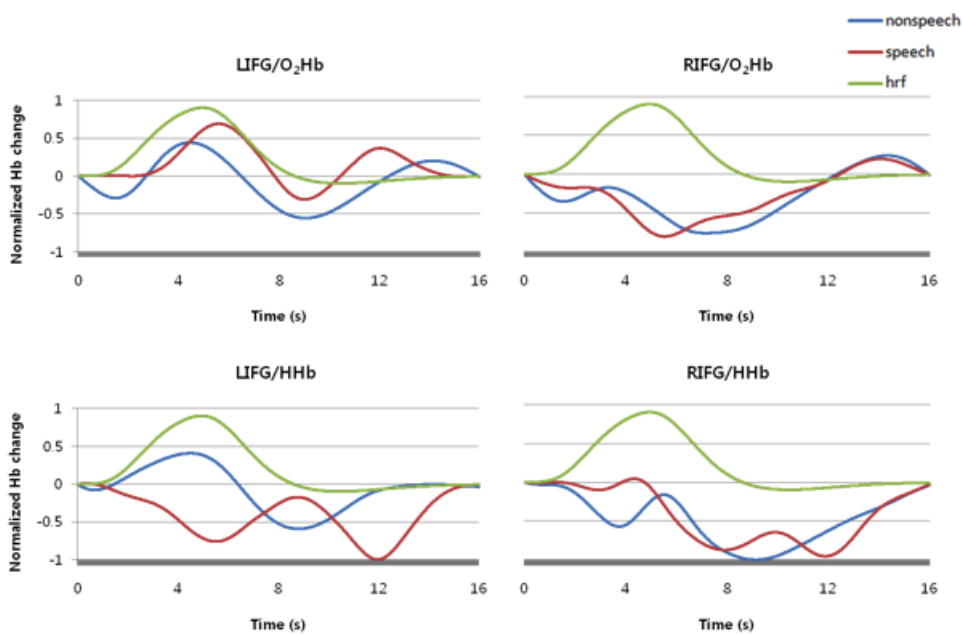


Figure 13. Hemodynamic responses at inferior frontal gyri (BA47)

At bilateral inferior frontal gyri, hemodynamic responses during the tasks were depicted. LIFG, left inferior frontal gyrus; RIFG, right inferior frontal gyrus; O_2Hb, oxygenated hemoglobin; HHb, deoxygenated hemoglobin.

Overall, in a similar shape, the O_2Hb change of nonspeech sounds at the LIFG was lower than that of speech sounds across the whole trial period. Due to the negative peak near 1.5 seconds, the positive peak of nonspeech sounds was found at about 4.45 seconds after the stimulus onset, which was about 1.15 seconds before that of speech sounds.

After 6 seconds after the stimulus onset, the deoxygenated Hb change of speech and nonspeech was significantly decreased at the RIFG, similar to the O_2Hb change. It indicates that total Hb concentration at the RIFG was decreased for both speech and nonspeech sounds. However, at the LIFG, the HHb change was positive for nonspeech immediately after the stimulus onset, in contrast to the negative change for speech, indicating that total Hb concentration of speech was not so changed for speech, whereas that of nonspeech was significantly increased (see Figure 14).

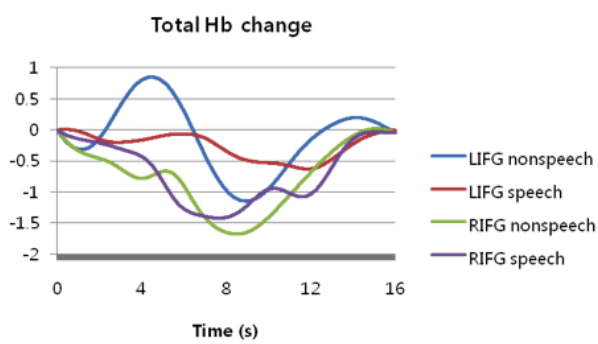


Figure 14. Total hemoglobin change at bilateral inferior frontal gyri

At bilateral inferior frontal gyri, total hemoglobin change was depicted. LIFG, left inferior frontal gyrus; RIFG, right inferior frontal gyrus.

For statistical comparisons between speech and nonspeech sounds, we took a mean amplitude measure of the 2-4 and 5-7 seconds time windows, encompassing the positive peaks of O_2Hb change. With two factors, stimuli and optode sites, two-way analysis of variance (ANOVA) test was carried on. Across the stimuli, we found no difference in the O_2Hb change from 2 to 4 seconds after the stimulus onset ($F(1,14) = 0.01$, $p = 0.9039$), while significant difference was found in the O_2Hb change from 5 to 7 seconds at 95 % confidence level ($F(1,14) = 4.07$, $p = 0.0486$). The optode effects were not found in both 2-4 ($F(1,14) = 0.49$, $p = 0.4865$) and 5-7 time windows ($F(1,14) = 0.03$, $p = 0.864$), and the interaction between stimuli and optodes was not significant, too ($F(1,14) = 0$, $p = 0.9852$ for 2-4 seconds; $F(1,14) = 0.77$, $p = 0.3853$ for 5-7 seconds).

Next, we compared passive listening of nonspeech sounds, i.e. natural sounds, animal vocalizations, and human emotional sounds (see Figure 15). It is notable that there was a significant O_2Hb change at the LIFG even though the subjects listened to the stimuli without motoric responses or verbal repetition.

Specifically, at the LIFG, we found hemodynamic modulation by the sound types. As shown in Figure 15, there was a positive peak of O_2Hb change evoked by listening to human emotional sounds at about 4.11 seconds after the stimulus onset. In case of natural sounds, a slightly lower positive peak was found at about 4.67 seconds after the stimulus onset, and a much lower positive peak was at about 5.08 seconds for animal vocalizations. At the RIFG, the positive peak of O_2Hb change by listening to human emotional sounds was consistently found, but only the negative peaks were found for the other two sounds.

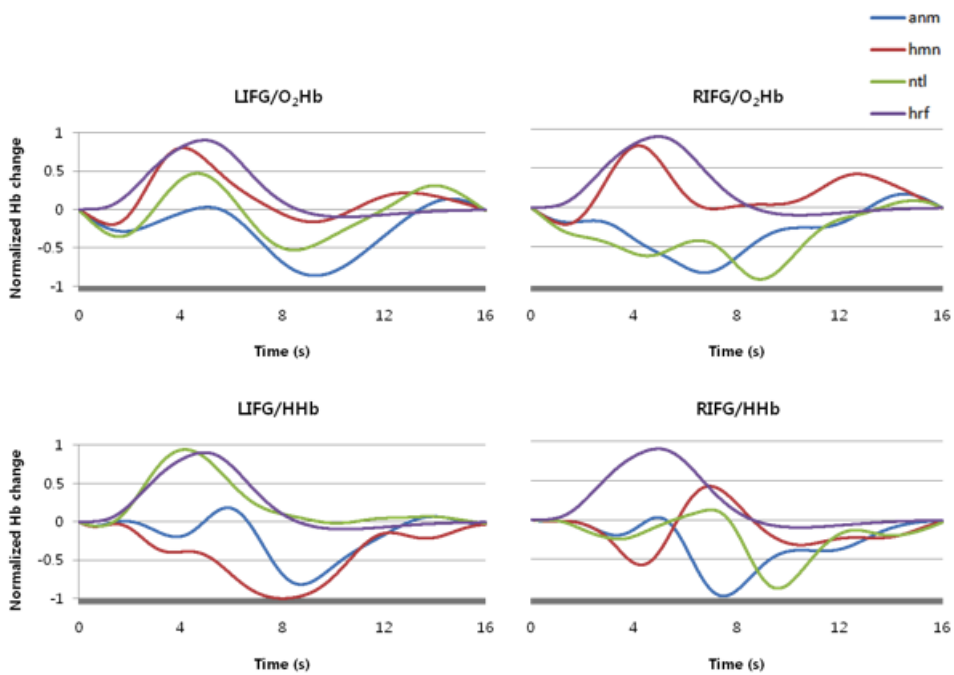


Figure 15. Hemodynamic responses of nonspeech sounds

At bilateral inferior frontal gyri, hemodynamic responses of passive listening of nonspeech sounds were depicted. LIFG, left inferior frontal gyrus; RIFG, right inferior frontal gyrus; O_2Hb, oxygenated hemoglobin; HHb, deoxygenated hemoglobin; anm, animal vocalizations; hmn, human emotional sounds; ntl, natural sounds.

However, statistical analyses with two-way ANOVA showed that the mean amplitude measure of the 3-5 seconds time windows after the stimulus onset was not significantly different among three nonspeech sounds at 95 % confidence level ($F(2,14) = 3.41$, $p = 0.068$). It might be probably because individual variance was much higher than the modulation effects. The optode position effects and the interaction between sound types and optodes were not significant, too ($F(2,14) = 1.04$, $p = 0.3575$ for optodes; $F(2,14) = 0.23$, $p = 0.7949$ for interaction).

2. Verbal repetition of words and pseudowords

We compared verbal repetition of words and pseudowords by calculating the proportion of O_2Hb change in total Hb concentration, as a functional index to indicate neural activities. Contrasting verbal repetition of words and pseudowords, we found that the O_2Hb changes of word repetition were higher than those of pseudoword repetition during the time windows of 4-7 seconds after the stimulus onset (see Figure 16). The result was statistically significant at 95 % confidence level ($F(1,14) = 5.95$, $p = 0.0162$; at 4-7 seconds time windows). Besides, it is more similar to the canonical HRF at the LIFG. The result implies that there were large blood supplies with O_2Hb to compensate O_2 consumption by neural activities. In contrast, we found no such change at the RIFG for both words and pseudowords.

For nonspeech sounds, there was no O_2Hb change at both LIFG and RIFG except for human emotional sounds showing small positive changes bilaterally (see Figure 16). It might be because total Hb concentration was simultaneously increased for nonspeech sounds as the O_2Hb increased. Statistical analyses using ANOVA revealed no main effects of sound types and optode positions, and no interaction between two factors was found at 95 % confidence level for the same time windows of 4-7 seconds after the stimulus onset ($F(2,14) = 1.13$, $p = 0.2901$ for sound types; $F(2,14) = 0.6$, $p = 0.5508$ for optodes; $F(2,14) = 0.6$, $p = 0.5494$ for interaction).

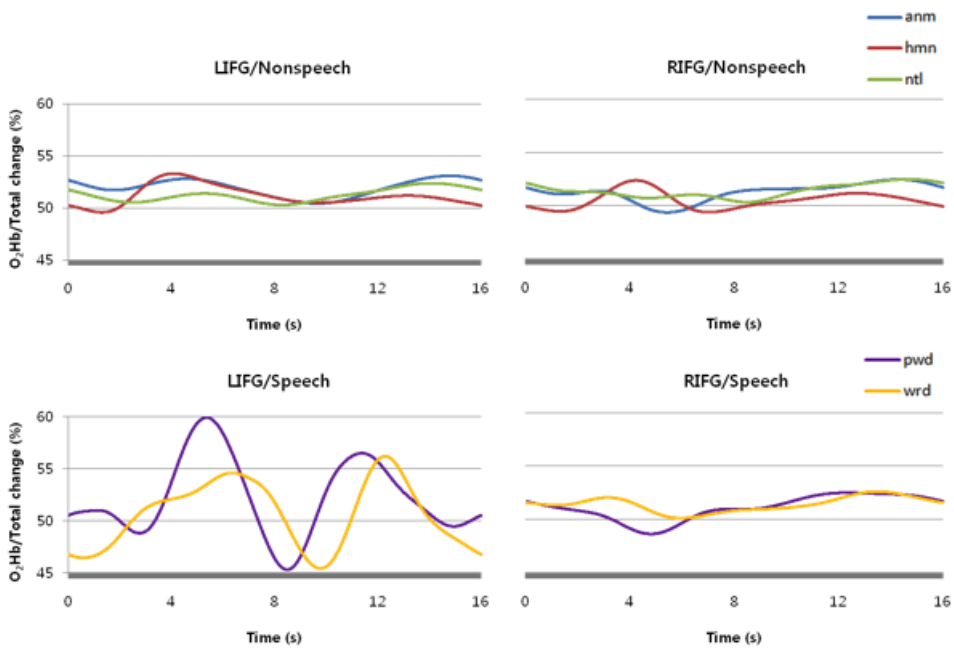


Figure 16. Percent change of O_2Hb in total Hb

At bilateral inferior frontal gyri, the percent change of O_2Hb in total Hb concentration by verbal repetition of words and pseudowords were shown. LIFG, left inferior frontal gyrus; RIFG, right inferior frontal gyrus; anm, animal vocalizations; hmn, human emotional sounds; ntl, natural sounds; wrd, words; pwd, pseudowords.

3. Systolic vs. diastolic pulsation and BOLD changes

Because of low temporal resolution, fMRI cannot discriminate pulsation by blood pressure from BOLD signals. fNIRS has much higher temporal resolution than fMRI and here, we tried to figure out how systolic and diastolic pulsations contribute to the hemodynamic change evoked by verbal repetition of words and pseudowords. To this end, we filtered out noises from raw fNIRS signals (cutoff frequency = 50 Hz) and localized positive and negative peaks (see Figure 17). At

individual peaks, we resampled the signals and extracted systolic and diastolic signals (see Data acquisition and analysis section).

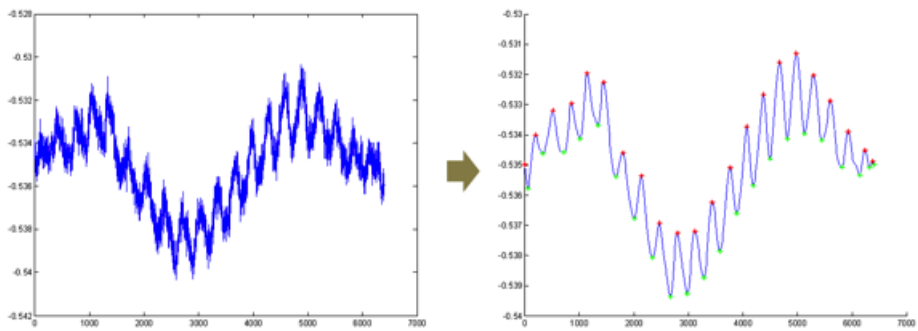


Figure 17. Localizing peaks on fNIRS signals

The signals were cut off at 50Hz and a sliding window algorithm (window width = sampling rate/2, slide distance = window width/2) was used to detect both positive and negative peaks (red, positive peaks; green, negative peaks).

With the reconstructed systolic and diastolic pulsations, we compared the O_2Hb/HHb changes evoked by verbal repetition of words and pseudowords. The results were shown in Figure 18.

For both words and pseudowords, the O_2Hb and HHb changes at the RIFG were negligible and there were little fluctuations by systolic/diastolic pulsations, either. At the LIFG, however, we found many fluctuations by systolic and diastolic pulsations for both stimuli. It implies that neural activities require sudden blood supplies locally, accompanying with large Hb fluctuations by systolic and diastolic pulsations. Overall, the patterns of systolic and diastolic changes were similar to that of the mean O_2Hb/HHb changes, but there were some divergent points between systolic and diastolic pulsations.

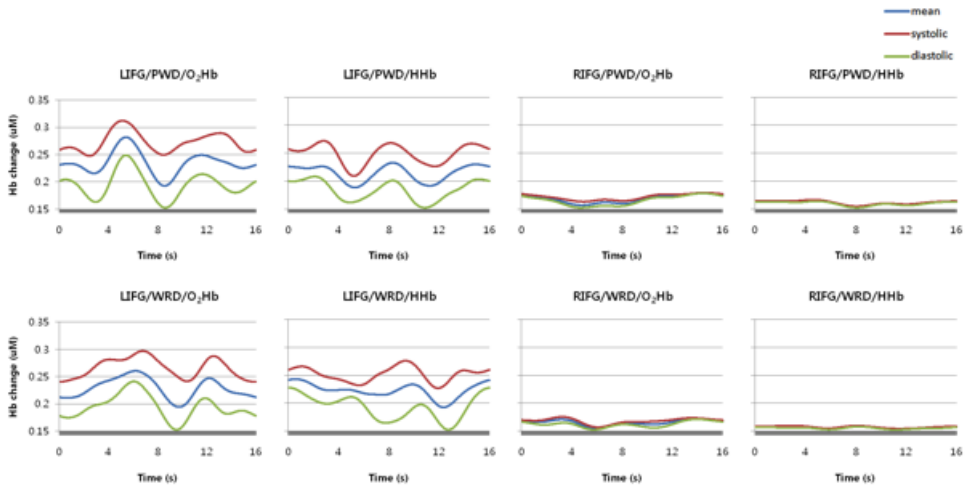


Figure 18. O_2Hb change according to systolic and diastolic pulsations

At bilateral inferior frontal gyri, in case of word and pseudoword repetition, the O_2Hb/HHb changes by systolic and diastolic pulsations were shown. LIFG, left inferior frontal gyrus; RIFG, right inferior frontal gyrus; WRD, words; PWD, pseudowords.

To investigate the characteristics of both pulsations in more detail, we additionally examined the percent change of O_2Hb in total Hb concentration for both systolic and diastolic pulsations. The results were shown in Figure 19. It is evident that the diastolic pulsation kept successfully track of mean O_2Hb change, though both systolic and diastolic pulsations showed similar patterns in terms of the proportions of O_2Hb in the total Hb concentration. In general, the systolic pulsation reflects cardiac output pressure, whereas the diastolic pulsation is determined by metabolic changes such as CO_2 concentration and neuronal activities at cardiac relaxation time. Therefore, it seems that the O_2Hb change by neuronal activities was mainly determined by the O_2 overcompensation occurring at diastolic phase.

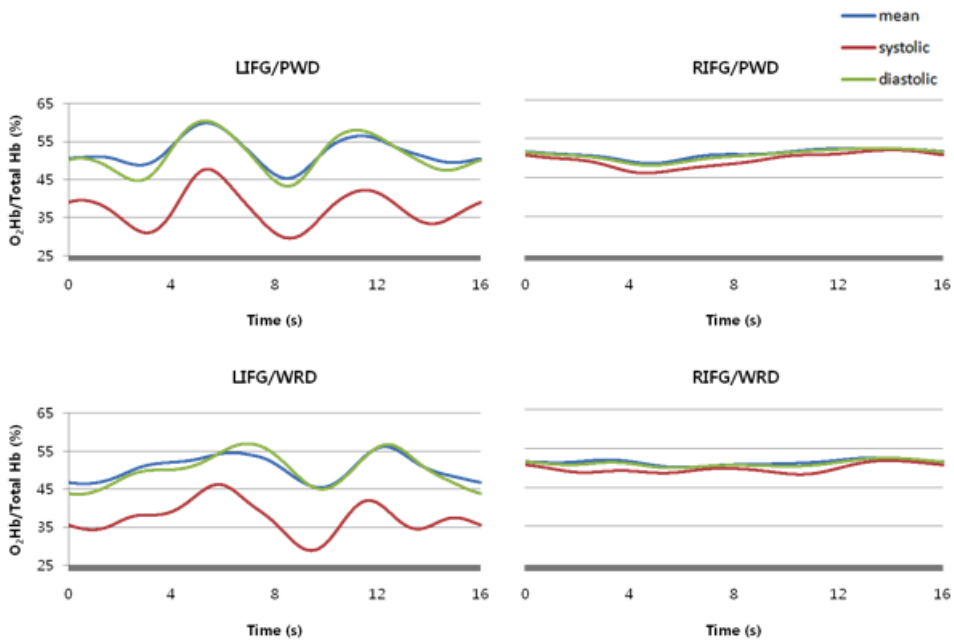


Figure 19. Percent change of O_2Hb in total Hb at systolic and diastolic phases

For systolic and diastolic pulsations, at bilateral inferior frontal gyri, the percent change of O_2Hb in total Hb concentration by verbal repetition of words and pseudowords were shown. LIFG, left inferior frontal gyrus; RIFG, right inferior frontal gyrus; wrd, words; pwd, pseudowords.

Last, we plotted phase plots indicating the phase difference between HHb and O_2Hb as a function of time (see Figure 20). As shown, there was weak linear relationship between HHb and O_2Hb in case of pseudowords at the LIFG, but not for words. In contrast, at the RIFG, we found that three phase plots for mean, systolic, and diastolic phase converged on an attractor within small region.

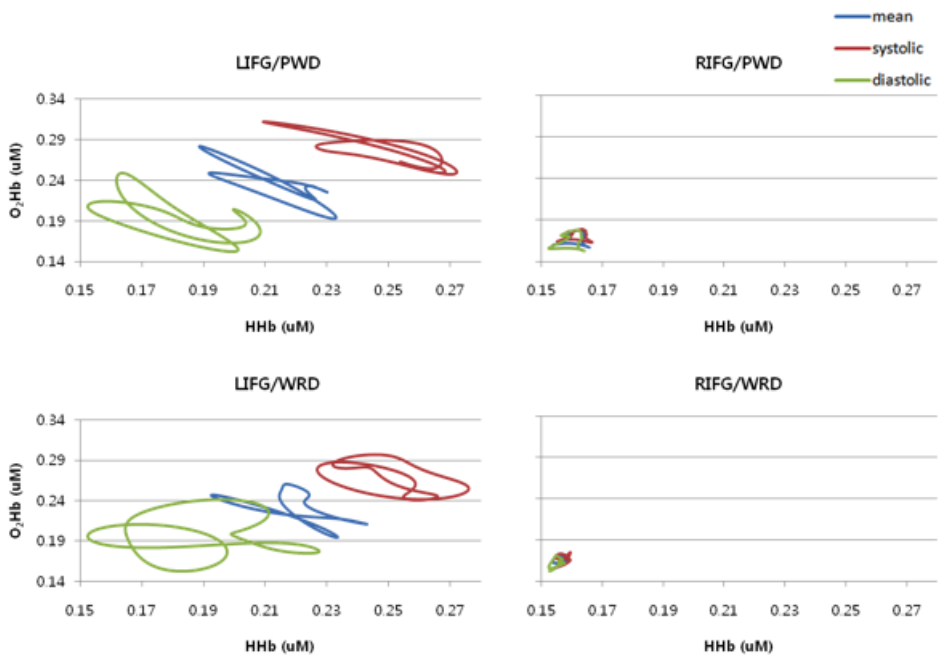


Figure 20. Phase plots between HHb and O_2Hb

Phase plots were depicted by positioning HHb and O_2Hb as a function of time, left inferior frontal gyrus; RIFG, right inferior frontal gyrus; wrd, words; pwd, pseudowords.

4. Discussion

We monitored the hemodynamic response at bilateral inferior frontal gyri (IFG, BA47) while the subjects simply listen to speech and nonspeech sounds, and verbally repeated speech sounds only. According to the results, we found there were significant hemodynamic changes at IFG even by passive listening of nonspeech sounds. Furthermore, we observed higher hemodynamic changes at LIFG by pseudoword repetition than those by word repetition, which is consistent with the previous findings in the first fMRI experiment.

1. Articulation-based sound perception

Speech perception has been traditionally considered in a sensory domain. Recently, however, some theories based on non-sensory domain are emerging to account for neural mechanism of speech perception. For example, the motor theory suggests that listener perceives not the acoustic features, but the abstract intended gestures required to articulate the sounds (Liberman and Mattingly, 1985). As another variant of the motor theory, direct realism tries to account for speech perception as perceiving actual vocal tract gestures using information in the acoustic signal (Fowler, 1986). These all presuppose that perceiving sounds intrinsically involves motoric movements (Fadiga et al., 2002).

After Broca's seminal discovery, the left inferior frontal gyrus (LIFG) was reported as the center of speech production of fluent, articulated speech as well as that of speech comprehension simultaneously (Caramazza and Zurif, 1976). This means that speech perception is partly dependent on the LIFG. In this vein, our results further suggest that the LIFG might be a center of perceiving nonspeech sounds as well as speech sounds. The nonspeech sounds such as natural sounds and animal vocalizations used in this study were not articulable in terms of human vocal organs. It is thus less likely that the subjects might subvocally articulate the nonspeech sounds while passive listening. Nonetheless, there were significant hemodynamic changes by perception of nonspeech sounds at the LIFG, which was comparable to speech sounds (see Figure 13).

With regards to this finding, it is notable that stimulus expectancy can modulate inferior frontal gyrus in passive auditory perception (Osnes et al., 2012). It is still debatable whether the LIFG has an essential or simple modulatory role in auditory perception. Nevertheless, motoric involvement is likely to be important in top-down control of auditory perception such as emotional arousal (Scott et al., 2010). This notion is additionally supported by various sensorimotor integration mechanisms (Pulvermüller et al., 2006; Wilson et al., 2004; Wilson and Iacoboni, 2006). Furthermore, neural activities at the LIFG can predict individual differences in perceptual learning of cochlear-implant patients (Eisner et al., 2010), indicating that learning of sound perception is partly dependent on the LIFG.

However, it is difficult to account for the hemodynamic modulation by the sound types at the LIFG. It might reflect the degree of internally simulated articulation to perceive the sounds, but it is not clear. Among them, it should be noted that human emotional sounds uniquely modulated the O_2Hb change at bilateral IFG, unlike the other sounds. With regard to this finding, Hoekert and colleagues revealed that left and right inferior frontal gyri were both involved in the processing of emotional prosody in speech, by measuring reaction time on emotional prosody task using rTMS (Hoekert et al., 2010). Another study with patients in supranuclear palsy reported that patients with gray matter atrophy in the RIFG showed significant correlations with voice emotion recognition and theory of mind deficits, indicating this region is associated with prosodic auditory emotion recognition (Ghosh et al., 2012). In other words, the bilateral change in

O_2Hb concentration by listening to human emotional sounds seem to be partly due to emotional process in speech perception.

Putting all together, an auditory-motor integration seems to be developed in parallel with cognitive demands to organize sounds as perceptually meaningful elements (Westerman and Miranda, 2002; Kuhl, 2004). It is also essential in social communication transferring nonverbal emotional states of others (Warren et al., 2006). Therefore, the hemodynamic changes at the LIFG suggest that auditory perception is in part supported by motoric representation, which is corresponding to articulation-based sound perception, proposed in the first experiment.

2. Articulatory representation of speech sounds

The main purpose of the second experiment was to re-examine the findings in the first experiment, i.e. whether pseudowords are differentially represented in the LIFG, compared to words. As discussed in the first experiment, we suppose that unfamiliar speech sounds such as pseudowords might use articulatory codes based on sound imitation at the LIFG and this is not the case in words. Therefore, we expected that the O_2Hb change was significantly higher for pseudowords than for words at the LIFG and the result was conformed to this expectation. It implies that the LIFG were likely to be reserved as a temporal storage of speech codes for pseudowords during verbal repetition (Yoo et al., 2012).

By the way, it is notable that there were relatively small but considerable increases in O_2Hb concentration by word repetition at the LIFG, too. It means that articulatory coding was automatically initiated by perceiving words at the LIFG. Unfortunately, due to the limitation of fNIRS channels, we could not measure the O_2Hb change at left middle temporal gyrus (LMTG), supposed to be a center of acoustic-phonetic codes of words. According to our previous results, however, it is more likely that the acoustic-phonetic codes at the LMTG became superior to the articulation-based codes at the LIFG for words. In this vein, two distinct neural activities at the LIFG and LMTG seem to be simultaneously evoked for perceiving words.

This might be in part because the LIFG serves as speech parser to detect word segmentation in continuous speech sounds (McNealy et al., 2006). McNealy and colleagues observed left-lateralized signal increases in temporal cortices only when parsing the continuous sounds containing statistical regularities, which was a precursor of words. More importantly, they found that neural activities at the LIFG and LMFG were positively correlated with an implicit detection of word boundaries, i.e. the detection of speech cues. In other words, the LIFG might act as speech segmentation circuits automatically recruited before auditory lexical retrieval was completed at the LMTG (Marslen-Wilson, 1987).

On the other hand, the LIFG was known as a part of human mirror neuron system, which was supposed to be based on imitation mechanism (Iacoboni, 2005; Iacoboni and Dapretto, 2006). It is consistent with the notion of articulation-based sound perception discussed in this experiment, in that unfamiliar sounds are likely

to be imitated for verbal repetition. In the same context, the O_2Hb change of word repetition observed in 4-7 time windows at the LIFG seems to originate from the analysis-by-synthesis facility to perceive the incoming speech sounds (Cohen et al., 1988).

Another thing to note here is the second positive peak observed in words and pseudoword commonly (see Figure 16). The peak was found at about 11.39 seconds after the stimulus onset in pseudoword repetition, followed by that of word repetition at about 12.28 seconds after the stimulus onset. The second peak seems to reflect the speech production after listening to the sounds. Consistent with the notion, no second peaks were found in nonspeech sounds because the subjects passively listened to nonspeech sounds without verbal repetition. The phase difference of second peaks between words and pseudowords might be due to the difference of preceding events for perception.

Last, it is interesting that no peaks were found in the proportions of O_2Hb change in total Hb concentration for nonspeech sounds even though there were significant O_2Hb changes at the LIFG (see Figure 15). It indicates that the change of total Hb concentration is pivotal for neural activities and there exists a strongly nonlinear relationship between neural activity and hemodynamic response, which remains to be seen in future study.

3. BOLD signal and Systolic vs. Diastolic pulsation

Neural activity change is coupled to blood oxygenation level dependent (BOLD) signals measured by fMRI. The neurovascular coupling is a collective term to indirectly refer hemodynamic responses that consist of changes in blood flow, blood volume, blood oxygenation, and so on (Kwong et al., 1992; Ogawa et al., 1992). In order to overcome this non-specificity of fMRI measure, we need to investigate brain hemodynamics directly as a supplementary measure.

As one of such, the O_2Hb change by systolic/diastolic pulsations at local brain regions is considerable. In the present study, we found that there was little fluctuation by systolic and diastolic pulsations at the RIFG, but a large fluctuation was found at the LIFG, in which neuronal activities were found. Notably, the O_2Hb change by mean signals was very similar to that by diastolic pulsation, whereas the O_2Hb change by systolic pulsation was significantly lower than that by mean signals though the overall fluctuation is similar. Therefore, it is likely that the O_2Hb changes evoked by local neuronal activities were mainly dependent on those at diastolic phase.

This can be used as an alternative measure for BOLD signals monitored by fMRI. As known, fMRI measures functional hyperaemia, i.e. the increase in blood flow evoked by neuronal activity brings in excess oxygenated blood. However, distinguishing effects by change in blood flow from that in cognitive state is task-dependent, rather than global and task-independent (Abel et al., 2003). There is also possibility that BOLD signals reflect astrocyte signaling that regulates cerebral blood flow to power neural computation as well as functional hyperaemia evoked by neuronal activities (Attwell et al., 2010).

In this case, the O_2Hb changes by systolic/diastolic pulsations clearly can distinguish BOLD changes by neuronal activities from those by cerebral blood flow (see Figure 19). The phase plots also show clear difference between LIFG and RIFG, indicating neuronal activities significantly may change relationship between systolic and diastolic phases (see Figure 20). It will need additional researches to build exact model of BOLD and systolic/diastolic pulsation.

Chapter 4. Associating meanings with sounds

In the first experiment, we found that acoustic sounds may be transformed into different speech codes according to subjects' perception even when the same sound waves are processed in human brain. This implies that by learning, speech sounds can be differentially represented in connection to specific concepts in long-term memory. Then, the next question will be how the learning is achieved in terms of neural plasticity, i.e. which brain regions are recruited for the learning? To answer this question, we aimed at examining brain regions mediating words learning – from novel sounds to known words. Specifically, we compared neural activities before and after learning while the subjects verbally repeat physically identical sounds. The result showed brain network associating concepts with specific sounds.

1. How can sounds be associated with a specific meaning?

How a novel sound comes to have specific meanings in human brain has been an intriguing topic in the literature of cognitive linguistics and neurolinguistics for a long time. New word learning often requires a considerable and effortful time until it is incorporated into human language system, i.e. the mental lexicon. A recent study revealed that learning of novel spoken words

required at least a day of consolidation by which neural representation of newly learned words can be strengthened and thus recognized more rapidly as existing words in the mental lexicon (Davis et al., 2008). With respect to the learning process, neuropsychological data suggest hippocampal structures as essential neural substrates of memory consolidation (Gooding et al., 2000). This process is largely supported by neural plasticity in cortical circuits that support long-term linguistic knowledge and rules to maintain the human language system at all linguistic levels (Gow, 2012; Jacquemot et al., 2003; Kuhl and Rivera-Gaxiola, 2008; Scott et al., 2000; Zhang and Wang, 2007).

To be learned successfully, sounds should be first represented and processed on auditory cortex and several corresponding regions, e.g. premotor cortex, inferior frontal cortex, and inferior parietal lobule, which are reciprocally mediated by each other (Rauschecker and Scott, 2009). Notably, there is lots of neuroscientific evidence demonstrating that once learned speech is differentially processed in specific brain regions (Belin et al., 2000; Binder et al., 2000; Newman and Twieg, 2001; Dehaene-Lambertz et al., 2005). It means that learning process might reorganize some brain regions to distinguish learned sounds from others. This learning effect was so evident that word-specific neural activities were repeatedly described in many neuroimaging studies. However, there is no complete description of how such learning is processed and what neural circuits are involved in learning.

It has been suggested that word learning is mediated by the phonological loop (PL) suggested in verbal working memory (Baddeley et al., 1998). However,

there is no direct evidence yet that the PL is an essential component for word learning. In contrast, neuropsychological data show that verbal working memory and some linguistic abilities, e.g. speech perception and sentence comprehension, can be dissociated from each other (Friedrich et al., 1984, 1985; Martin, 2006). This raises the question whether word learning is independently processed from normal speech circuitries. All these confusion might be from the fact that the notion of working memory is likely to be confounded with higher mental processes such as cognitive aptitudes (Cowan et al. 2005, 2006).

In the third experiment, we aimed to investigate how novel sounds come to have specific meanings and what neural mechanisms mediate it. For this reason, we introduced verbal repetition task to show what neural circuits are recruited for learning process from sounds to meanings. The verbal repetition task is very suitable for studying speech processing in that it is simple, natural, and dynamic (Yoo et al., 2012). Subjects repeated novel sounds and then were asked to learn some of them by reading two short stories. They could easily learn the novel sounds by associating the sounds with specific meanings. After the short learning, subjects were again asked to repeat the same sounds. During the repetition tasks, we monitored brain activities by functional MRI in event-related design as used in the first experiment.

2. Experimental Design

1. Subjects and Stimuli

Nineteen native Korean adults (6 males and 13 females) aged 18-26 years old (mean 20.2 years) participated voluntarily in the present study. Informed consent was obtained from all participants before the experiment. All participants had normal auditory ability and reported no neurological deficits. The subjects completed a questionnaire that assessed their handedness, according to the Edinburgh Handedness Inventory (Oldfield, 1971), and all were strongly right-handed (scored 80 or higher). The experiment was conducted according to protocols approved by the Institutional Review Board of Gachon University of Medicine and Science.

We used a minimal set of pseudowords to provide with the appropriate learning conditions and ensure the mastery learning occurred (Bloom, 1968). As a result, only ten pseudowords were generated by randomly combining several consonants and vowels in Korean. Each pseudoword consisted of three syllables to balance the syllable lengths between stimuli and had no meanings in Yonsei Korean Corpus 1-9 (Yonsei Korean Dictionary, 1998). Five out of ten pseudowords were presented to subjects as unlearned stimuli while the others were presented as learned stimuli during verbal repetition (see Experimental Procedures for detail descriptions). The pseudowords spoken by a male Korean native speaker were recorded and converted into computer files of *.wav* format (22,050 Hz, 16 bit, stereo). The loudness (average RMS level) of all stimuli was normalized (-60 to 0 dB) by a sound software (SoundForge; Sony Creative Software Inc.), and thus it

was not significantly different between the stimuli. All stimuli were aligned with the onset of the first syllable and did not exceed two seconds in total length.

Two different stories, a condensed story based on Rapunzel (German fairy tale collected by the Brothers Grimm) and a newly created story (a short diary written by an imaginary boy) for this study, were used for the learning of pseudowords. We presented them visually to the subjects in random order. The stories did not exceed 180 words (Rapunzel: 179, Diary: 163), and thus they could be presented on a single screen without any difficulty in reading (Font size: 18). For learning, each story contained five pseudowords respectively. All pseudowords were highlighted by different colors to make them salient. Importantly, the meaning of pseudowords was naturally explained in one story, but not in the other one, leading to the classification of learned and unlearned stimuli used in the verbal repetition. The learned and unlearned stimuli set were counterbalanced between the subjects.

2. Experimental Procedure

There were two fMRI scanning sessions and one learning session outside the MRI scanner. Two scanning sessions, in which the subjects were asked to listen binaurally and repeat overtly what they heard (simple verbal repetition task), included four different runs respectively. To focus on the task, we asked the subjects to repeat the stimuli at approximately the same speed and keep their eyes closed during the scanning session. One run consisted of twenty-one trials

with two seconds of listening, another two seconds of repeating, and twelve seconds of resting, and thus was totally 5 min 48 sec (see Figure 21). The selected pseudowords were randomly presented in one run and the subjects had totally four runs in a random order. Since a pseudoword was presented at least twice in one run, all pseudowords were presented at least 8 times during one scanning session. Two scanning sessions were separated as scans before and after learning by the learning session.

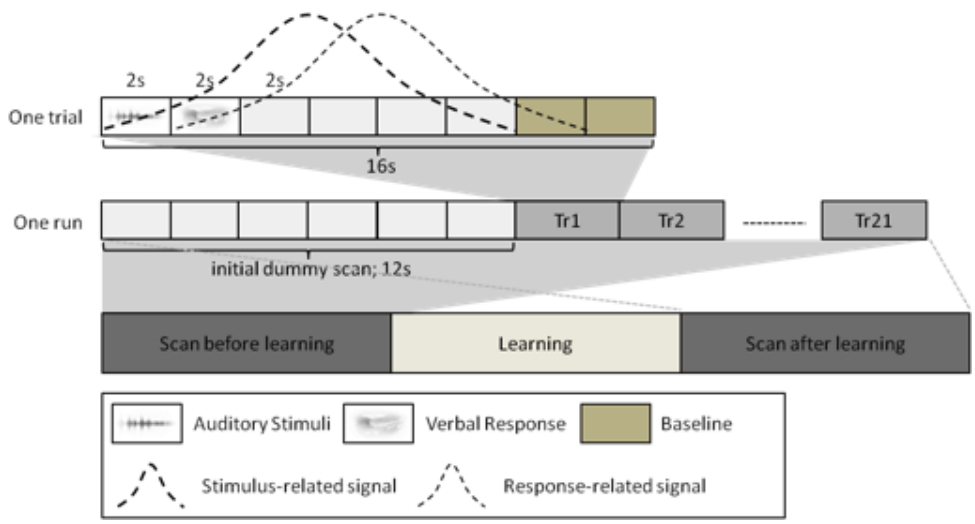
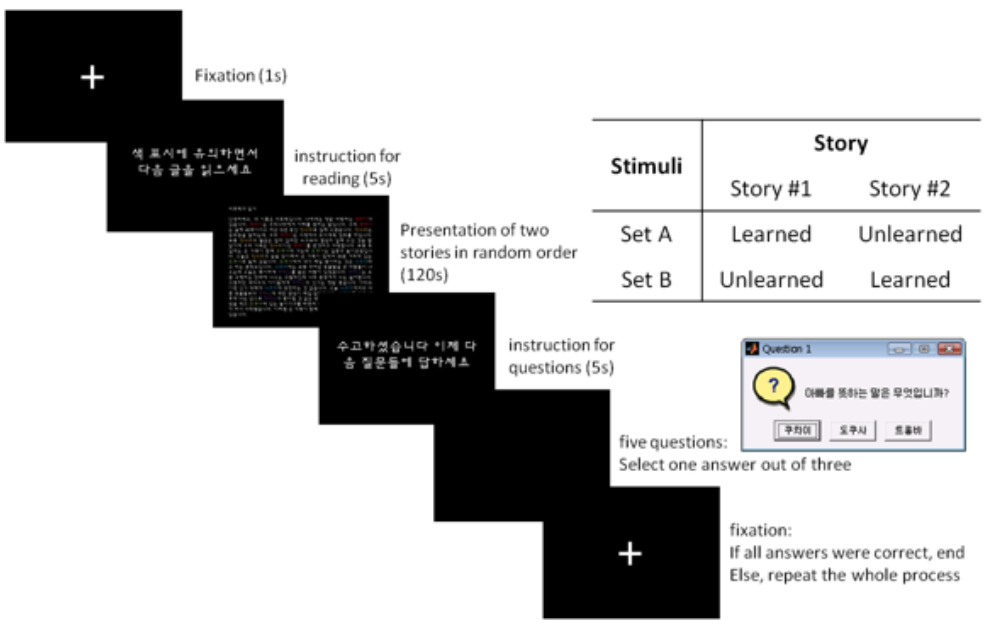


Figure 21. Experiment design for the third experiment

Two separate scans were conducted for each subject, each of which was before and after learning respectively. One scanning session consisted of four runs and one run included twenty-one trials. One trial is 16 seconds in length: 2 seconds of perception, 2 seconds of production, and last 12 seconds of resting. There was an initial 12 seconds for signal stabilization of fMRI.

After the scan before learning was done, the subjects were asked to read two stories containing the selected pseudowords outside the scanner according to the pre-informed protocol (see Figure 22). We did not explicitly ask them to memorize the pseudowords and their meanings while reading the stories. After a

short fixation (1 sec) was presented, for five seconds, an instruction was given for the subjects to ask them to read carefully the following stories. Then, two different stories were shown on the whole computer screen (LG LCD Monitor L1753T, 17 inch) for 120 seconds respectively. The monitor screen was about 50 cm away from the eyes. The reading speed of the average adult is at 250 to 300 words per minute (wpm) in case of prose text, and while proofreading materials, people usually read at 200 wpm on paper and 180 wpm on a monitor (Ziefle, 1998). Therefore, the presentation time of the stories was enough to read and understand them. The presentation order of two stories and the stimuli set were randomly selected for learned and unlearned stimuli.



Stimuli	Story	
	Story #1	Story #2
Set A	Learned	Unlearned
Set B	Unlearned	Learned

Figure 22. Learning procedure for verbal repetition of acoustic sounds

Learning procedure was conducted outside the scanner. Two stories were presented to the subjects; one contained five pseudowords as learned stimuli and the other five pseudowords as unlearned stimuli. Verifying whether the subjects learned pseudowords was done in self-paced manner. See the text for detail descriptions of the whole procedure.

The learning of pseudowords was assessed after reading the stories. The meanings of learned pseudowords were self-explained in the stories, each of which was asked by five consecutive questions requiring one answer out of three in a self-pace manner. The subjects should answer each question by a button press. If all answers were correct, the learning session was closed. If not, however, the whole learning session re-started from the beginning. For most subjects, the learning session was finished within two iterations and did not exceed three times even at the worst case. The average duration of learning session was 10.2 min. All instructions, stories, and questions were sequentially displayed on the same computer screen (see Figure 22). After the learning session, another scanning session (scan after learning) was launched, which was the same to the first one, but with different presentation order of stimuli.

3. Data acquisition and analysis

While lying in an MRI system with a 12-channel head coil (3T Verio, Siemens Medical Solutions), the subjects conducted the repetition tasks. For all runs in the experiment, the imaging protocol consisted of isotropic 3-D T_1 MPRAGE and BOLD fMRI sequences with echo-planar imaging (EPI). T_1-weighted anatomical images were obtained first (TR = 380 msec, TE = 3.06 msec, FA = 70°). Then T_2*-weighted EPI images were acquired at the same slice locations of the T_1 image, parallel to the anterior commissure-posterior commissure (AC-PC) line, with the following parameters: brain volumes = 132 (n = 132), TR = 2000 msec, TE

= 30 msec, FA = 90°, FOV = 220 x 220 mm^2, matrix size = 64 x 64 pixels, slice thickness = 3.5 mm (with 3.5 mm gap), number of slices = 30. Each period began with a control image acquisition to ensure the label image was acquired during neuronal activity and to limit synchronization problems between the respective beginnings of the acquisition and thus total acquisition time was 5 min 51 sec. For this reason, the six images for the first twelve seconds were discarded to better approximate a steady state in the MR signal.

The auditory stimuli were presented inside the MRI system via a magnetically shielded audio system (SS-3100 Silent Scan, Avotec Inc.), but it is in general hard for subjects to listen and repeat exactly what they heard in a noisy MRI system. To prevent the scanner noise from interfering with auditory perception and verbal repetition, here we employed the interleaved silent steady state (ISSS) sequence, a sparse imaging method that allows brief silent periods for verbal repetition between image-acquisition pulses (Schwarzbauer et al., 2006). In an event-related design, the subjects could listen and repeat the incoming sounds during silent period (4 seconds), and the neural activities evoked by verbal repetition were successfully measured during subsequent imaging period (12 seconds) (see Figure 21). To minimize head movements, little foam cushions were used to wedge next to subjects' heads and the subjects were asked to articulate as clearly as possible in a quiet voice.

SPM5 (Wellcome Trust Centre for Neuroimaging, UCL) was used for preprocessing of functional brain images. First, the images were aligned to the first volume of the corresponding sequence using affine transformation with six

parameters (translation and rotation). According to Jezzard and Clare (1999), 2-3 pixels of local distortion in the EPI images can result from a 5° flexion. Thus, we excluded data with distortion greater than 2° flexion (corresponding to about one pixel or 3.5 mm) from the analysis, leading to minimization of the speech artifact problem in spite of the overt speech. Second, slice timing of the realigned images was corrected to match the different timing of MR signals between the first and last slices. Then, to be sure that all images were in the same coordinates, the brain volumes were spatially normalized to the Montreal Neurological Institute (MNI) template that covers the whole brain. Last, Gaussian filtering with 8 mm full width at half maximum (FWHM) was applied to smoothen the images.

After preprocessing of brain volumes, statistical analysis was performed for the whole brain according to the general linear model (GLM) (Frackowiak et al., 1997). Neural activities were modeled by canonical hemodynamic response functions (HRFs) at every trial onset time. In a rapid event-related design of short inter-stimulus interval (ISI), a jittering is in general used to randomize the onset of stimuli and enhance the design efficiency of functional MRI (Dale, 1999). However, we used a fixed ISI to make two subsequent events, perception and production, to be equally processed in all trials, which made it hard to separate HRFs of both events due to severe co-linearity problem between two consecutive events. To avoid such problems, therefore, we modeled both perception and production as a single event in this study.

We investigated which brain regions were consistently activated by verbal repetition tasks, irrespective of the types of stimuli. Then we contrasted two

conditions with learned and unlearned stimuli in terms of activated volumes and loci. More importantly, we investigated regional correlation between activated loci, which were assumed to be neural correlates of the mirror neuron system in human (Iacoboni, 2005; Iacoboni and Dapretto, 2006). For Further analysis, we used dynamic causal models (DCM) provided in SPM toolbox to investigate how those regions were related to each other and what roles they have in verbal repetition (Friston et al., 2003; Daunizeau et al., 2010). The region-of-interests (ROIs) for DCM analysis were selected from activated loci in the first and third experiments: left inferior frontal gyrus (BA47), left middle temporal gyrus (BA39), bilateral superior frontal gyri (BA9, 10), and bilateral inferior parietal lobules (BA7, 40).

The results were summarized and analyzed in terms of word learning and mirror neuron system. All neural activities were identified in individual subject analysis and then statistically modeled in the group analysis by considering random effects. The results (*t* statistical maps of significantly activated voxels) were mapped on the MNI-template brain and rendered on 3-D brain with the help of SPM5 and MRICro (Rorden and Brett, 2000). For visualization of activated regions, the MNI coordinates were converted into the Talairach coordinates by a nonlinear transform to get a good match for both the temporal lobes and the top of the brain in all images (Talairach and Tournous, 1988).

3. Results

1. Neural activities before and after learning

We first mapped task-related neural activities on the template brain. During repetition of novel sounds, after pooling all four conditions, we found neural activities in the following regions (green area; see Figure 23): left superior temporal gyrus (BA22), left middle temporal gyrus (BA21), left cerebellum (declive), left middle frontal gyrus (BA10), left superior frontal gyrus (BA9), left cerebellum (uvula), left postcentral gyrus (BA3), left precentral gyrus (BA4), right superior temporal gyrus (BA38), right middle frontal gyrus (BA10), right insula (BA13), right lentiform nucleus (putamen), right inferior frontal gyrus (BA47), right cerebellum (tuber), right cerebellum (declive), and right middle frontal gyrus (BA46). Overall, the task-related neural activities were found frontal and temporal areas bilaterally, around the perisylvian region and premotor areas, and there were also significant activities in paracentral lobule and precuneus.

Next, we compared neural activities in four different conditions, where the stimulus types and learning phases were differentiated by two-by-two design. When subjects repeat unlearned stimuli, the meanings of which were not self-explained in the presented story, we found that superior and middle frontal gyri (BA9, 10) were bilaterally activated only after learning condition (white circles; see Figure 23a).

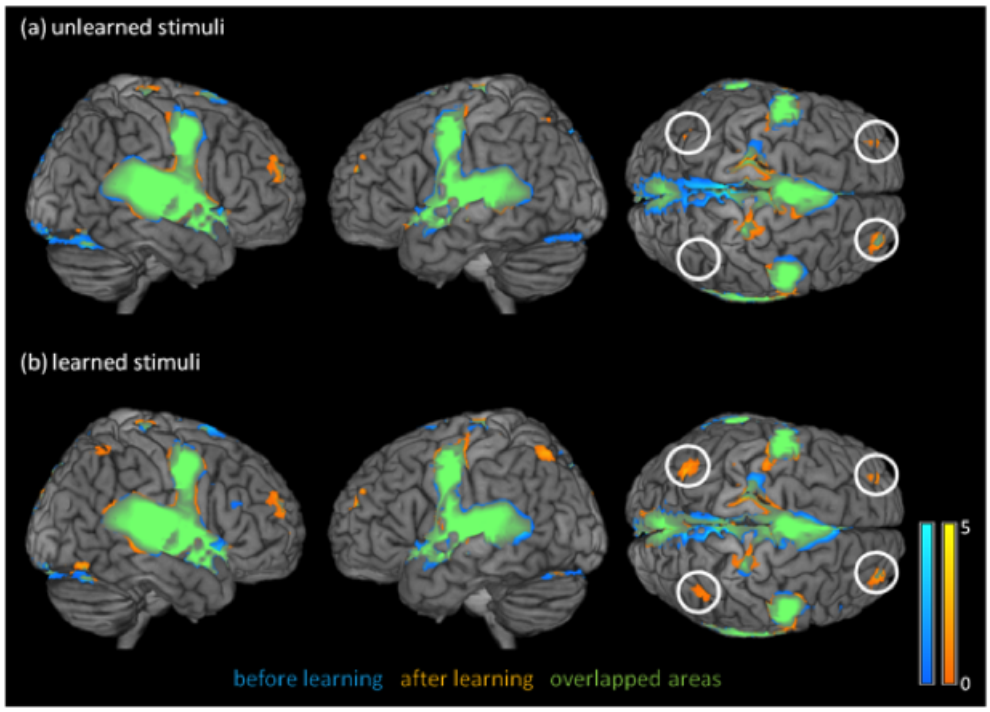


Figure 23. Neural activities before and after learning

Neural activities before and after learning were compared to reveal neural changes after learning ($p < 0.05$, corrected). The neural activities before learning were indicated as blue, while the neural activities after learning were orange-yellow. The green areas indicate the overlapped regions between before and after learning.

In case of repeating learned stimuli, in addition to superior and middle frontal gyri, superior and inferior parietal lobules were bilaterally activated after learning condition (white circles; see Figure 23b). The activated loci were summarized in Table 4. Notably, the superior and inferior parietal lobules were selectively activated for verbal repetition of learned stimuli.

Table 4. Neural activities newly evoked after learning

Conditions	Brain region	Cluster size	*t*-value	*z*-value	*x,y,z* (mm in MNI)
Unlearned stimuli after learning	L. Superior/Middle Frontal Gyrus (BA9, 10)	25	4.16	3.44	(-32, 52, 24)
	R. Superior/Middle Frontal Gyrus	35	3.75	3.18	(32, 56, 20)

	(BA9, 10)				
Learned stimuli after learning	L. Superior/Inferior Parietal Lobule (BA7)	21	3.51	3.02	(-36, -64, 56)
	R. Superior/Inferior Parietal Lobule (BA40)	11	2.72	2.45	(44, -56, 56)
	L. Superior/Middle Frontal Gyrus (BA9, 10)	29	3.96	3.32	(-32, 52, 24)
	R. Superior/Middle Frontal Gyrus (BA9, 10)	27	3.28	2.86	(32, 56, 20)

Only activations with a $p < 0.05$ (corrected) and a volume of at least 640 mm^3 (10 measured voxels) were considered. The x, y, and z values show the center of gravity of the activated clusters in Montreal Neurological Institute (MNI) coordinates. L, left; R, right; BA, Brodmann area.

It is intriguing that a few loci in the frontal and parietal areas were uniquely activated after learning, not observed before learning. In the line of the experiment 1, however, it is expected that inferior frontal gyrus (BA47) and middle temporal gyrus (BA39) in left hemisphere were also activated after learning as a result of articulatory and semantic learning, respectively. Thus, we further analyzed the result within small region-of-interest (ROI) areas: BA44, 45, 47 (for inferior frontal gyrus) and BA39 (for middle temporal gyrus (BA39). The masked ROI image was automatically generated by WFU_pickatlas software (Maldjian et al., 2003), based on the Talairach Daemon database (Lancaster et al., 1997; Lancaster et al., 2000). The result was listed in Table 5.

Table 5. Neural activities newly evoked after learning (only in masked areas; BA44,45,47,39)

Conditions	Brain region (Masked with BA44,45,47,39)	Cluster size	*t*-value	*z*-value	*x,y,z* (mm in MNI)
Unlearned stimuli after learning	L. Inferior Frontal Gyrus (BA47)	10	3.21	2.82	(-52, 20, -8)
Learned stimuli after learning	L. Superior Occipital Gyrus (BA39)	21	7.68	1.75	(-36, -76, 24)
	L. Angular Gyrus (BA39)		5.20	1.57	(-44, -68, 36)
	L. Middle Temporal Gyrus (BA39)	31	7.18	1.72	(-48, -72, 16)
	L. Middle Temporal Gyrus		7.15	1.71	(-48, -60, 24)

	(BA39)			
	L. Middle Temporal Gyrus (BA39)	5.92	1.63	(-35, -56, 24)

Only activations with a $p < 0.05$ (corrected) and a volume of at least 640 mm^3 (10 measured voxels) were considered. The x, y, and z values show the center of gravity of the activated clusters in Montreal Neurological Institute (MNI) coordinates. L, left; R, right; BA, Brodmann area.

As shown in Figure 24b and Figure 24c, by masking ROI areas only, there were unique neural activities by verbal repetition of unlearned stimuli in left inferior frontal gyrus (BA47) after learning, while there were unique neural activities by verbal repetition of learned stimuli in left angular gyrus (BA39) and middle temporal gyrus (BA39) after learning. It is considerably overlapped with the activated loci (white circles, see Figure 24a) observed in the first experiment.

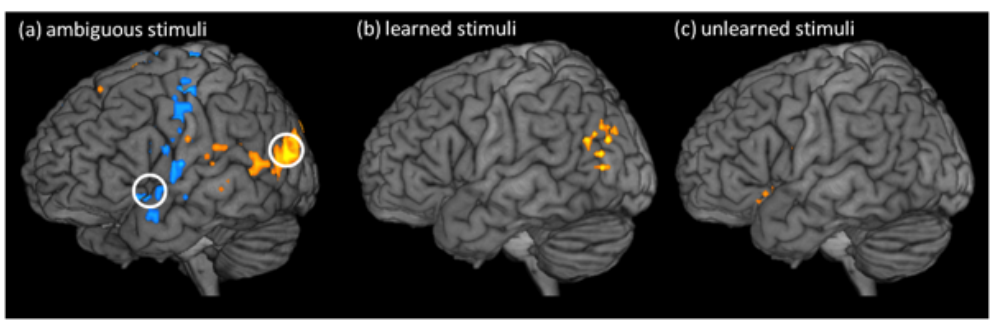


Figure 24. Neural activities after masking ROI areas

(a) Neural activities identified in the first experiment: orange-yellow (word-perceived) and blue (pseudoword-perceived) (b) Neural activities after learning for learned stimuli. (c) Neural activities after learning for unlearned stimuli. (b) and (c) are all obtained in masked ROIs (see the text). ($p < 0.05$, corrected).

Last, we investigated which brain regions were not activated after learning, i.e. we mapped unique neural activities found only before learning and not observed after learning. The results were shown in Figure 25 and Table 6.

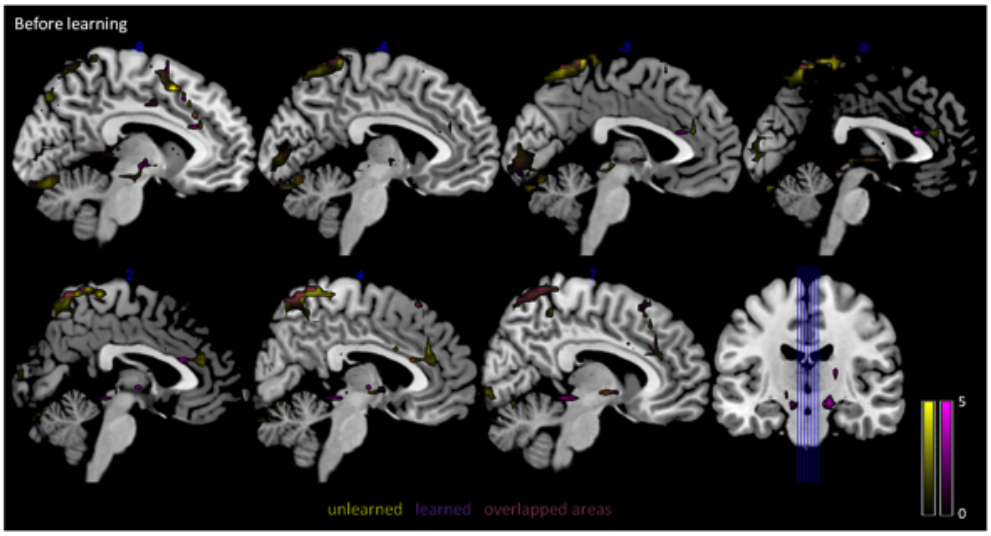


Figure 25. Brain regions uniquely activated before learning

Neural activities within some medial parts of bilateral brain showed negative neural changes after learning (p < 0.05, corrected). The neural activities before learning in unlearned stimuli were indicated as yellow, while the neural activities before learning in learned stimuli were indicated as violet. The violet-red areas indicate the overlapped regions between unlearned and learned stimuli.

For both unlearned and learned stimuli, the neural activities were mainly found in medial parts of the brain: bilateral insula (BA47), right putamen, left supplementary motor area, and a few parts of left cerebellum were found in unlearned stimuli, whereas right insula and thalamus, left anterior cingulate, right posterior cingulate, left supplementary motor area, and right precuneus were found in learned stimuli. Overall, no significant difference was found between unlearned and learned stimuli in terms of activated clusters.

Table 6. Local maxima of brain regions uniquely activated before learning

Conditions	Brain region	Cluster size	*t*-value	*z*-value	*x,y,z* (mm in MNI)
Unlearned stimuli before learning	L. Cerebellum (Declive)	509	4.99	1.55	(-32, -64, -24)
	L. Limbic Lobe (Uncus)	9	3.41	1.36	(-32, 4, -24)
	R. Lentiform Nucleus (Putamen)	147	6.59	1.68	(16, 8, -8)
	R. Insula (BA47)	31	3.76	1.41	(36, 20, 0)

	L. Insula (BA47)	12	2.63	1.23	(-36, 20, 0)
	L. Supplementary Motor Area	87	5.18	1.56	(-8, 12, 52)
Learned stimuli before learning	R. Cerebellum (Declive)	185	4.00	1.44	(32, -64, -24)
	R. Insula (BA47)	58	4.04	1.44	(36, 20, 0)
	R. Thalamus	182	5.85	1.62	(12, -8, 0)
	R. Posterior Cingulate	8	2.76	1.25	(4, -40, 20)
	L. Anterior Cingulate	67	5.94	1.63	(0, 24, 20)
	L. Supplementary Motor Area	20	4.13	1.45	(-8, 16, 52)
	R. Precuneus	69	4.90	1.54	(4, -48, 72)

Only activations with a $p < 0.05$ (corrected) and a volume of at least 640 mm^3 (10 measured voxels) were considered. The x, y, and z values show the center of gravity of the activated clusters in Montreal Neurological Institute (MNI) coordinates. L, left; R, right; BA, Brodmann area.

2. Regional correlations between activated loci

We identified two significant brain regions that were likely to be recruited for meaningful sound learning. We further analyzed how those regions were correlated with each other to obtain learned speech sounds. To this end, we first investigated regional correlations between activated loci such as superior frontal gyrus and inferior parietal lobule bilaterally, together with left inferior frontal gyrus and left middle temporal gyrus that were identified in the first experiment.

We obtained % BOLD change at those regions as shown in Figure 26. With these time-series data, we calculated autocorrelation and cross-correlation between the selected ROIs to figure out the signal similarity (or coherence) between them.

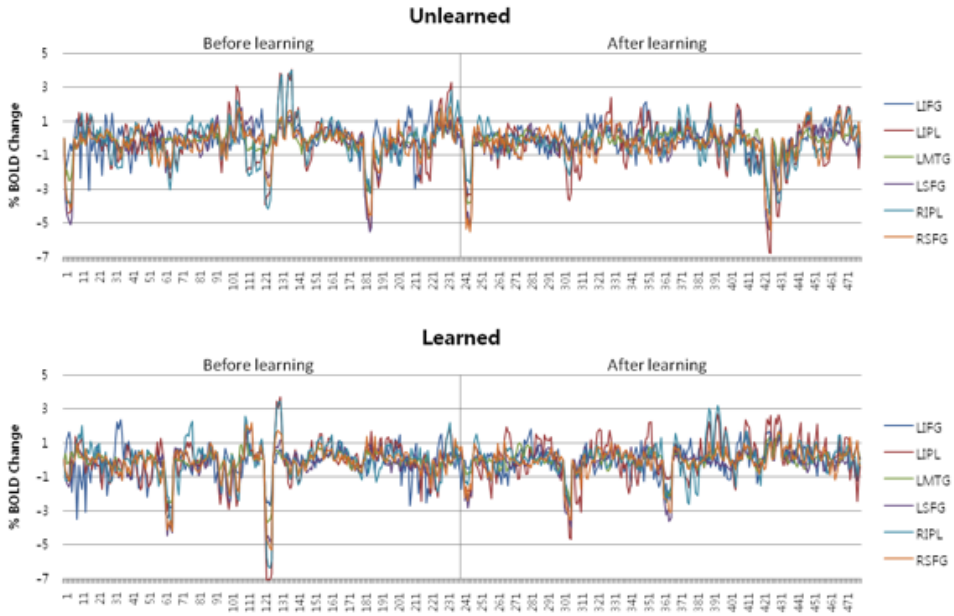


Figure 26. % BOLD change at six ROIs

For unlearned stimuli, the autocorrelation of each locus was shown in Figure 27. As indicated with red arrows, the autocorrelation at LIFG was enhanced after learning, compared to the one before learning. In contrast, the autocorrelation at LMTG was slightly diminished after learning. It is consistent with the finding that LIFG was more activated in case of pseudoword-perceived verbal repetition in the first experiment.

It is also notable that the autocorrelations at LSFG and RSFG were smoothed at several peaks around the center point, indicating that neural activities became more complex after learning. At left and right IPLs, the autocorrelations were notably enhanced after learning.

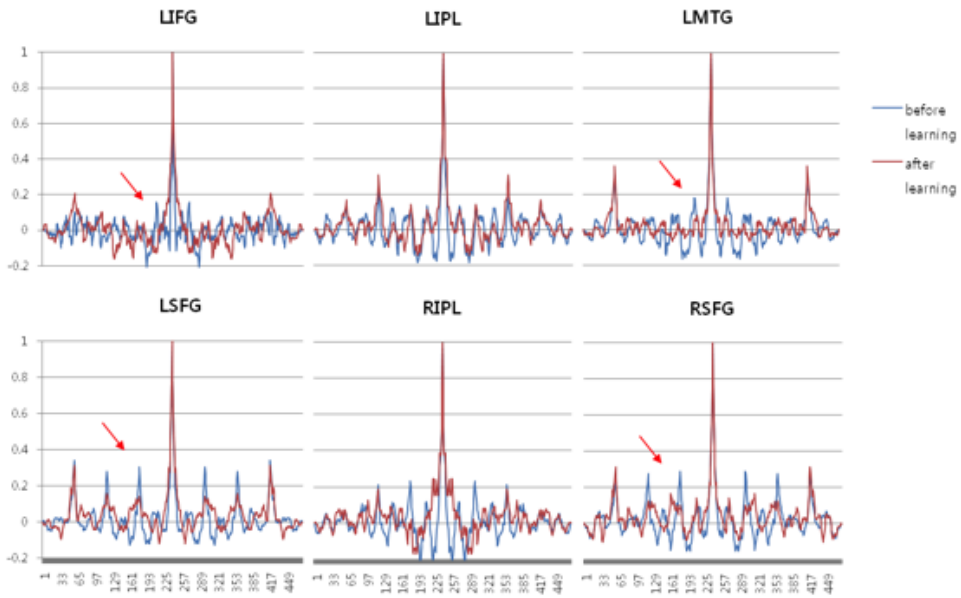


Figure 27. Autocorrelations before and after learning at six ROIs (unlearned stimuli)

Interestingly, in case of learned stimuli, the autocorrelation at LIFG was slightly diminished at the center point after learning, together with several peaks emerged around the center point, which was not the case at LMTG (red arrows; see Figure 28). More intriguing was found in left and right SFG, in which the autocorrelations were clearly diminished after learning, indicating that neural activities at those regions were suppressed after learning. In LIPL and RIPL, the autocorrelations were also slightly diminished at the center point. All these findings imply that there might be a regional connectivity for speech processing before and after learning respectively, located in frontal and parietal areas.

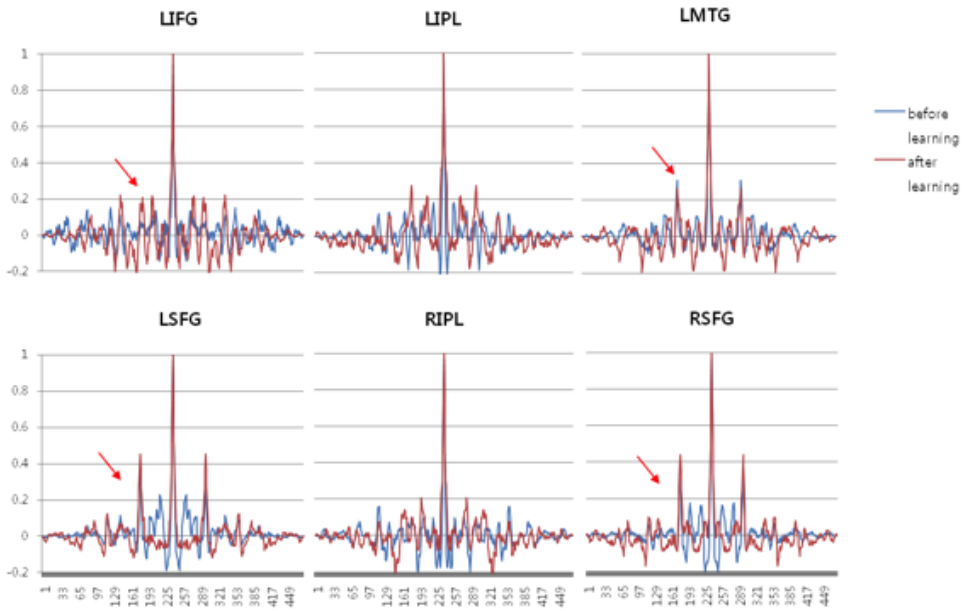


Figure 28. Autocorrelations before and after learning at six ROIs (learned stimuli)

The cross-correlations between ROIs support the notion of regional connectivity. For example, there was enhanced cross-correlation between LIPL and RIPL after learning in case of unlearned stimuli (see Figure 29). No significant difference was found in LIFG-LSFG and LIPL-LMTG pairs. However, in case of learned stimuli, the cross-correlation between LIPL and RIPL after learning was not so significant (see Figure 30). Instead, there were enhancement of correlation between LIP and LMTG and diminishment of correlation between LIFG and LSFG. Therefore, it is likely that by learning, some regions have stronger connectivity while the others weaker.

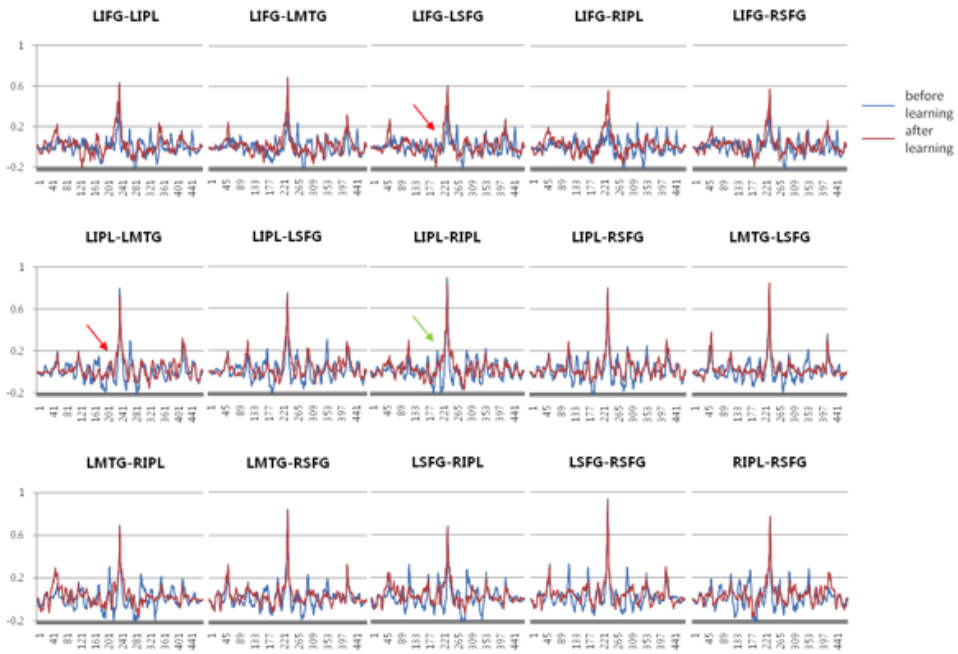


Figure 29. Cross-correlation between ROIs in unlearned stimuli

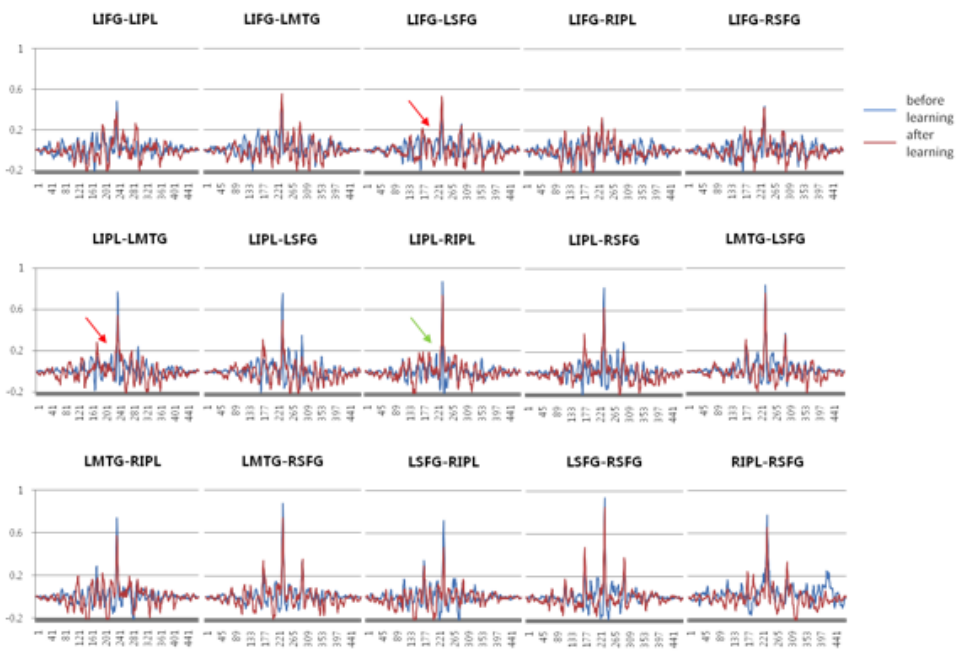


Figure 30. Cross-correlation between ROIs in learned stimuli

3. Dynamic causal models of word learning

The regional correlations provided an idea of neural networks mediating learning from sounds to speech. However, the cross-correlation is not enough to show causality between those regions. Thus, we further analyzed this regional connectivity by using dynamic causal modeling provided in SPM toolbox. We presupposed several models accounting for the activated patterns, in which the unlearned stimuli required only articulatory learning whereas the learned stimuli required both articulatory learning and semantic learning simultaneously. According to activated loci, we also assumed that the articulatory learning was mediated by superior frontal gyri and the semantic learning was mediated by inferior parietal lobules. Then, we compared goodness of fit (GOF) of the models according to Bayesian model selection (BMS).

The compared models and three evaluation indexes (AIC, BIC, and F) were listed in Table 7. As shown, full connection (FC) model, in which all the loci were connected to each other reciprocally, was not good to account for neural activities after learning, indicating that there should be some specific connectivity between the loci (see Figure 31a). Reduced connection (RC) model assumed only some links between neighboring loci and its GOF was much better than the FC model (see Figure 31b).

Table 7. Bayesian Model Selection for word learning

Model	FC	RC	AC	VC	DC	DOC	DDC
AIC	-1984.64	-1984.18	-1971.54	-1981.83	-1983.63	-2019.22	-2028.63

BIC	-2160.8	-2125.6	-2100.31	-2123.33	-2115.13	-2140.79	-2145.25
F	-17726.9	-17739	-17727.9	-17741.4	-17740.4	-17755.7	-17760.9

Based on the RC model, we assumed simple connectivity between six loci as indicated in Figure 31c, called assumed connection (AC) model. However, the GOF of AC model was very low, similar to the FC model. In the first experiment, by the way, we discussed the cooperative division in speech processing supported by Perisylvian connectivity revealed by DT-MRI (Catani et al., 2005). This notion led us to compare two models based on ventral- and dorsal-pathways, respectively.

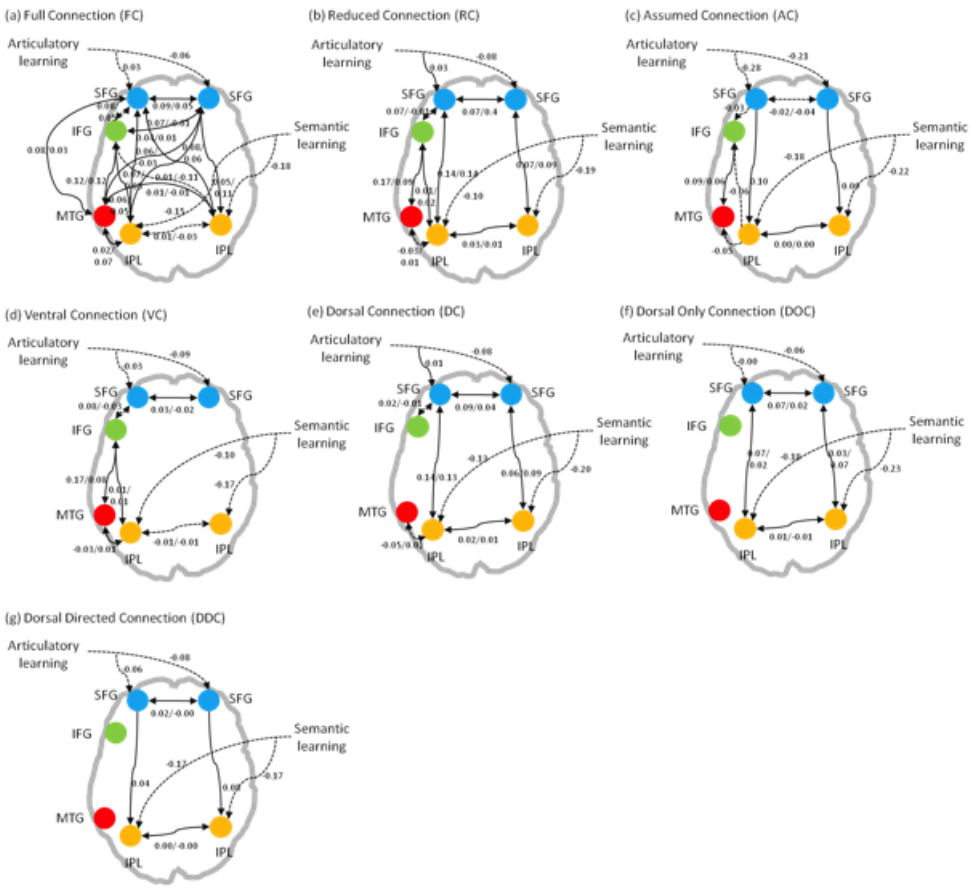


Figure 31. Suggested DCM models with intrinsic connectivity

The ventral connection (VC) and dorsal connection (DC) models were shown in Figure 31d and Figure 31e. The VC and DC models showed better GOF than the previous models, similar to each other. Then, we disconnected the links between activated loci in the first experiment (LIFG and LMTG) and activated loci in the third experiment (SFG and IPL), leading to dorsal only connection (DOC) model (see Figure 31f). The DOC model showed had better GOF than VC and DC, indicating that the activated loci in the first experiment might have little role in word learning. Last, we investigated whether there was a directional connection between the dorsal loci. We assumed directional connections from SFG to IPL in dorsal directed connection (DDC) model (see Figure 31g). The result showed the best GOF in the compared models, indicating that word learning was likely to be mediated by dorsal network with a directional connectivity from SFG to IPL bilaterally.

4. Discussion

We investigated which brain regions were involved to learn meanings of novel sounds in the second experiment. In verbal repetition of novel sounds after associative learning, we observed two specific neural areas, i.e. superior/middle frontal gyrus (BA9, 10) and superior/inferior parietal lobules (BA7, 40) were newly activated after learning. In contrast, we found that medial parts of the brain, e.g.

insula (BA47), putamen, thalamus, anterior/posterior cingulate, supplementary motor area, and a few parts of left cerebellum were not activated after learning, indicating that these brain regions were not recruited for speech processing after learning. We also examined directional connectivity between these regions by DCM and found the fronto-parietal networks subserve as associative learning of novel sounds.

1. Neural circuits mediating associative learning

A novel auditory sound in this experiment should be acoustically analyzed and temporarily maintained in short-term memory for successful verbal repetition. Before learning, speech codes for the novel sounds were not stored in the mental lexicon. Therefore, repeating the sounds is similar to sound imitation that requires articulatory codes for temporary usage as revealed in the first experiment (Yoo et al., 2012). After learning, however, repeating the sounds might be divided into two different conditions. In the first condition, in which unlearned stimuli were verbally repeated, by learning phase, the subjects knew exactly how to articulate the sounds to repeat them even though they did not know the meanings of sounds. In the second condition with learned stimuli, they knew the meanings of sounds as well as how to articulate the sounds from the explanation given in the stories. In short, there might be two kinds of learning in this experiment.

An auditory scene with continuous acoustic waves should be parsed as recognizable phonetic and phonological items stored in the brain to be verbally

repeated. Such producible items or speech codes are usually formulated in the mental lexicon by language acquisition. Therefore, for unfamiliar sounds before learning, we should assimilate the sounds into some known items represented in our phonetic and phonological space to repeat them. However, the pseudowords in this experiment consisted of familiar phonetic and phonological items for the subjects. If so, for unlearned stimuli, what the subjects learned from the learning phase was likely to how to name the novel sounds.

For example, similar situations may occur in case of face recognition. We already have knowledge of individual features consisting of the face, i.e. eyes, ears, and nose, each of which has specific visual features to distinguish one from the others. However, how each feature is organized to be a face is more important to recognize the face. A novel face to us is not familiar even though each feature or item consisting of the face is familiar or known to us. In this sense, knowing or learning a novel face means that we can name the face as a whole by associating the visual scene with specific abstract representation to maintain the scene in the brain. Though the visual scene has a specific name to refer it, there is no semantic information about the scene yet before learning. That is, in addition to be exposed to the novel face, a different kind of learning is required to learn the face in detail.

The unlearned stimuli in this experiment were simply, repeatedly exposed to the subjects without explanation of their meanings, similar to the novel face recognition. In this case, the auditory scenes were novel to the subjects because they have never used phonemes and syllables to constitute the pseudowords in their native language. Therefore, they might need an abstract representation, i.e.

names to maintain the sounds in the brain. In this process, the pseudowords came to have associated names without semantic information. In contrast, by learning phase, the learned stimuli were not only repeatedly exposed to the subjects, but also given with relevant explanation of their meanings. We classified these two kinds of learning as articulatory learning and semantic learning in this study.

The notion of articulatory and semantic learning is in part supported by another study by Tsukiura et al. (2002). In the retrieval of newly learned people's names and occupations with Brain-damaged patients and fMRI, Tsukiura and his colleagues showed that bilateral prefrontal areas near to the activated loci by repeating the unlearned stimuli in this study, i.e. left superior frontal gyrus ([-24 63 12]) and right middle frontal gyrus ([35 56 18]) were crucial for the process of associating newly learned people's faces and names. In addition, they found that left superior parietal lobules ([-38 -66 54], [-42 -66 54]), which were near to the activated loci by repeating the learned stimuli in this study, were activated by contrasting novel and familiar stimuli both in names and occupations, indicating that the loci were recruited to the process of associating newly learned people's faces with semantic information (occupations) as well as names.

In sum, the associative learning appears to be divided into two stages: associating the object with its name and its semantic contents. It is accompanied with a cortical reorganization in bilateral prefrontal cortices and parietal lobules. The learning process also seems to be applicable for both visual and auditory scenes in common. If so, what neural mechanisms support the learning process? We will try to answer this question in the next discussion.

2. Associative learning and episodic buffer

It is known that dorsolateral prefrontal cortex is a crucial part of working memory and it may promote long-term memory formation by strengthening associations among items that are organized in the working memory (Blumenfeld and Ranganath, 2006). The prefrontal cortex (BA9) and temporo-parietal junction (BA40) are more sensitive to the cross-modal (auditory-visual) memory task, but not the memory load itself (Zhang et al., 2004). Besides, posterior parietal cortex (PPC) contributes to episodic memory retrieval modulated by relevant attention (Wagner et al., 2005; Hutchinson, Uncapher, and Wagner, 2009). All these are reminiscent of the episodic buffer that was supposed as a temporary storage of information from multimodal subsidiary systems (Baddeley, 2000).

Interestingly, the prefrontal cortex and temporo-parietal junction were overlapped with neural circuits reserved for mediating articulatory and semantic learning in this study. As discussed, the activated loci seem to be neural correlates of associative learning irrespective of the modality of objects. In this vein, it is notable that the prefrontal activities were likely to be associated with episodic long-term memory supporting the formation and retrieval of memories for events as well as working memory for maintenance and manipulation of information over short delays (Ranganath, Johnson, and D'Esposito, 2003). In general, learning of novel spoken words require at least a day of consolidation to be represented as existing words in the mental lexicon (Davis et al., 2008), which is mediated by

hippocampal structures (Gooding et al., 2000). That is, it is likely that encoding of information is dissociable from the information itself in the brain (Vargha-Khadem et al., 1997).

Putting all together, therefore, the prefrontal cortices and parietal lobules localized in this study were likely to act as temporary storages used before consolidation in the long-term memory. Before consolidation, multimodal sensory information is organized as a single event or object by strengthening associations among items organized in the episodic buffer, called as associative learning of multisensory information. In this vein, we can speculate the respective roles of prefrontal cortices and parietal lobules. Recently, Champod and Petrides found that the posterior parietal cortex (PPC, inferior parietal sulcus, [-38 -48 48]) and the mid-dorsolateral prefrontal cortex (MDLFC, [21 40 26]) have distinct roles in working memory (Champod and Petrides, 2007). According to them, the PPC is centrally involved in manipulation processes, whereas the MDLFC is related to the monitoring of the information that is being manipulated.

In the monitoring process, the semantic information of the objects is not required, while the same information is important in the manipulation process. It is consistent with our findings in that semantic learning in superior and inferior parietal lobules manipulates both auditory sounds and their semantic information, while articulatory learning in superior and middle frontal gyri only deals with the auditory sounds. As shown in Figure 31, it is likely that there is a functional connectivity from the prefrontal cortex to posterior parietal cortex, indicating that directional learning from articulation to meaning. The directionality of the fronto-

parietal networks was also confirmed in this study by the DCM analysis. This fronto-parietal network is likely to be modulated by attentional control (Wagner et al., 2005; Wang et al., 2009). In addition, its effective connectivity seems to be reduced in patients with working memory deficits due to schizophrenia (Deserno et al., 2012). In summary, the associative learning is likely to be based on this fronto-parietal network reserved for episodic working memory.

Chapter 5. General Discussion

1. Neuroanatomy of verbal repetition

Before being repeated, the incoming sounds are recoded, stored, and rehearsed, which formulates a phonological loop with short-term memory processes involved (Henson, Burgess, and Frith, 2000; Jacquemot and Scott, 2006; Paulesu, Frith, and Frackowiak, 1993; Vallar et al., 1997). However, the immediate repetition task used in the present study does not seem to involve a rehearsal phase precisely because it takes only a very short time to listen and repeat the sounds. The recoded speech sounds are likely to be directly transferred into a phonological output buffer via phonological short-term memory (Baddeley, 2003a, 2003b). In this case, speech codes are mainly dependent on the process of cycling sensorimotor information between input and output buffers recruited at the phonological level in perception and production (Howard and Nickels, 2005; Jacquemot and Scott, 2006). The same loop seems to be equally applicable for nonspeech stimuli such as musical sounds (Koelsch et al., 2009). Nevertheless, many studies suggest there are distinct neural circuits used to process sounds in different categories (Belin et al., 2000; Binder et al., 2000; Dehaene-Lambertz et al., 2005; Husain et al., 2006; Newman and Twieg, 2001; Scott et al., 2000). Consistent with this literature, we found that the repetition task recruited not

only a common auditory-motor interface but also distinct neural circuits selectively recruited by specific stimuli.

In the classical view, the neural activities observed in this study can be divided into two separate but interactive networks: frontal and temporal. The frontal network, localized in the bilateral premotor cortex (BA6), Broca's area (BA44, 45), and the latter's right homologue, corresponds to the articulation of target sounds. The prefrontal network, including the dorsolateral prefrontal cortex (BA46) and the inferior prefrontal gyrus (BA47), is likely to be correlated to syllabification (Hickok and Poeppel, 2004; Indefrey and Levelt, 2004). In contrast, the temporal network is involved in the processing of incoming sounds. They are first processed at Heschl's gyrus (BA41) and then spectrotemporally analyzed in the superior temporal gyrus (BA21, 22, 38, 41, 42) and middle temporal gyrus (BA21) (Hickok and Poeppel, 2007). In addition, the Sylvian-parietal-temporal (Spt) area is enacted by simultaneous involvement of speech perception and production (Binder et al., 2000; Hickok and Poeppel, 2004). It is to be noted that distinct neural activities modulated by subjects' perception are not easily incorporated into this classical view, as it requires an account of phonological and semantic interaction at the stage of low-level speech processing.

The notion of a phonological learning device can provide a possible account of our results (Baddeley, Gathercole, and Papagno, 1998). In many neurolinguistic models, the auditory-motor interface has been regarded as a linguistic device in general (Ben-Shalom and Poeppel, 2008; Grodzinsky and Friederici, 2006; Hickok and Poeppel, 2007; Indefrey and Levelt, 2004; Poeppel

and Hickok, 2004; Vigliocco, 2000). In this vein, the neural circuits recruited during repetition tasks can constitute a general learning device to process incoming sounds, with a cooperative division of internal modules grounded on the specific contents of those sounds (Chomsky, 2000; Cummings et al., 2006; Kemmerer et al., 2008; Pinker, 1994). For instance, an unfamiliar sound may be learned after several imitations by activating a phonological loop (Baddeley, 2002; Baddeley and Hitch, 1974). Once the sound becomes familiar, it is usually associated with certain meanings by learning faculties. In this way, an acoustically identical sound will be differentially processed after it has been learned. The dual-stream model suggests the existence of two different pathways with regard to the outcome of this learning process (Hickok and Poeppel, 2007). That is, two separate networks are involved in generating articulation-based and acoustic-phonetic codes, which correspond respectively to the pseudoword-specific and word-specific activities in this study. Perisylvian connectivity revealed by DT-MRI (diffusion-tensor magnetic resonance imaging) also supports the existence of a cooperative division in speech processing. Catani et al. (2005) reported a direct pathway between Wernicke's area and Broca's area active in fast, automatic word repetition and an indirect pathway where verbal comprehension and semantic/phonological transcoding intervene between verbal input and articulatory output. This implies that two pathways are selectively involved in speech processing. However, the division in the roles of the two pathways appears to vary by neural efficiency, as reported in the context of speech perception (Dehaene-Lambertz et al., 2005; Guenther et al., 2004).

2. Vocal imitation and auditory-motor interface

Vocal imitation is rare in animals, but some animals, e.g. songbirds and elephants are known to be able to do it (Schachner et al., 2009). In humans, vocal imitation or auditory-oral matching capabilities play an important role in speech language acquisition (Kuhl and Meltzoff, 1996; Chen et al., 2004). Furthermore, language change gradually occurs as a result of vocal imitation (Harrington, 2000). Therefore, vocal imitation is considerable as a promising tool to investigate how we can learn speech from novel sounds. However, vocal imitation in speech is not easy to study because of its complex dynamics accompanying online auditory-motor integration.

Instead, in the present study, we aimed at examining neural mechanism of verbal repetition. Verbal repetition basically originating from vocal imitation is a pivotal ability in intact language function and it is suitable for studying online speech processing as well as word learning (Corrigan, 1980; Iverson et al., 2003; Kuhl et al., 1992; Wallesch and Kertesz, 1993; Massaro and Cowan, 1993). The verbal repetition formulates a phonological loop equipped with verbal short-term memory (Jacquemot and Scott, 2006; Paulesu et al., 1993; Vallar et al., 1997), in which incoming speech sounds are to be recoded, stored, and rehearsed before repeated by listeners. Most of such processes were mainly dependent on a left-

lateralized network involving Broca's area, dorso-lateral premotor cortex, supra-marginal gyri, and posterior temporal regions in human brain (Henson et al., 2000).

In the first experiment, we contrasted the transient verbal information maintained in the phonological short-term memory. Since we adopted immediate repetition paradigm with little rehearsal process, the recoded speech sounds were likely to be directly transferred between input and output buffers recruited at the phonological level. In this way, we could investigate speech codes temporarily maintained through cycling sensorimotor information (Howard and Nickels, 2005; Jacquemot and Scott, 2006).

Another thing to be noted is that we focused on the perceptual difference in auditory sounds by introducing novel ambiguous sounds, whereas most studies were confounded with the acoustic difference found in auditory sounds (Belin et al., 2000; Binder et al., 2000; Dehaene-Lambertz et al., 2005; Husain et al., 2006; Newman and Twieg, 2001; Scott et al., 2000). In addition, the verbal repetition task was relatively free from higher-level cognitive processes, which enabled us to separate speech processing maintained in audition-articulation loop from others.

Consequently, we found that an auditory-motor interface was essential in verbal repetition, which was also corresponding to a phonological learning device (Baddeley et al., 1998). The same interface was equally applicable for nonspeech such as musical sounds (Koelsch et al., 2009), implying the relationship between verbal repetition and vocal imitation. That is, the auditory-motor interface is not only a linguistic device, but also a phonological learning device that is capable of handling novel auditory sounds.

For example, once a novel sound heard, we can articulate a novel sound by phonological imitation or assimilation, which is mediated in the auditory-motor interface (by articulatory learning), and then usually can associate it with a specific meaning (by semantic learning). That is, meaningful sounds are likely to be differentially processed in human brain after the learning (Cummings et al., 2006; Kemmerer et al., 2008). This is exactly what we found in this study: Without semantic learning, novel sounds would have articulation-based codes by verbal repetition. However, acoustic-phonetic codes at left middle temporal gyrus were predominant after semantic learning.

3. Neural mechanism of speech sounds learning

According to dual-stream model (Hickok and Poeppel, 2007), two distinct neural pathways were likely to be reserved for articulatory and semantic learning, respectively. In this vein, semantic and phonological transcoding was mediated by the auditory-motor interface between phonological input and articulatory output. This notion is not easily incorporated into the classical view of speech processing, in which perception and production are separated from each other. To account for this, we need to investigate how the learning is correlated with the auditory-motor interface because learning seems to have an important role in the process, leading to enhancement of neural efficiency in the context of speech perception (DeHaene-Lambertz et al., 2005; Guenther et al., 2004).

A novel sound consists of an unfamiliar auditory scene with continuous acoustic waves. To be understood, it should be parsed as recognizable phonetic and phonological items in the brain. In this process, two learning processes are required: one for naming the sounds, and the other for understanding the sounds. The former means that we need to build an abstract representation to refer the sounds because the auditory scene was novel to us and we have no corresponding speech codes in the mental lexicon. The latter indicates that the referred name of sounds is to be associated with specific semantic information, which provides us with relevant understanding of the sounds.

It is similar to the novel face recognition. We have knowledge of individual features consisting of the face, but have no idea of how a novel face is organized with these features. By seeing, we come to have an abstract representation of the face. That is, we can distinguish it from others by learning the visual scene. After this kind of learning, we can additionally associate it with specific meanings that are correlated with the visual features of the face. It corresponds to the second learning processed at semantic level.

In this vein, Tsukiura and his colleagues showed that bilateral prefrontal areas were crucial for the process of associating newly learned people's faces and names (Tsukiura et al., 2002), which was near to the activated loci by repeating the unlearned stimuli in the second experiment. In addition, they found that left superior parietal lobules were activated by contrasting novel and familiar stimuli both in names and occupations (Tsukiura et al., 2002), which was found in the repetition of learned stimuli.

Therefore, it is likely that leaning sounds requires two different steps: one for associating the auditory object with its name and the other for associating the name with its semantic contents. By the way, it is also intriguing that the learning process seems to be applicable for visual and auditory scenes in common. With regard to this, we noted that the activated loci found in the second experiment were significantly overlapped with the prefrontal cortex and temporo-parietal junction, supposed as neural correlates of episodic buffer (Blumenfeld and Ranganath, 2006; Zhang et al., 2004; Wagner et al., 2005; Hutchinson et al., 2009; Baddeley, 2000).

While responding to an episodic event, that is, learning an auditory or visual scene seems to be dealt with in the episodic buffer, irrespective of the modality of objects. That is, before consolidated, the episodic buffer supports the formation and retrieval of memories for events (Ranganath et al., 2003; Davis et al., 2008; Vargha-Khadem et al., 1997). As a temporary storage, the episodic buffer maintains and organizes multimodal sensory information as a single event by strengthening associations among individual items.

Probably, the respective roles of prefrontal cortices and parietal lobules in the episodic buffer were monitoring and manipulation of the auditory objects (Champod and Petrides, 2007). Naming the sounds does not require the semantic information (monitoring), but the same information is important in associating the sounds with specific meanings (manipulation). There is a directional connectivity from the prefrontal cortex to posterior parietal cortex, indicating manipulation is initiated after monitoring. This fronto-parietal network is likely to be modulated

by attentional control (Wagner et al., 2005; Wang et al., 2009; Deserno et al., 2012).

However, it is still unclear how speech codes are correlated with the associative learning. The connectivity analysis using DCM showed a directional connectivity within the episodic buffer, but at the same time, there was less connectivity between neural circuits for speech codes and associative learning. It implies that there might be another neural mechanism to support speech codes. The second experiment may shed lights on this puzzle.

4. Sound perception and sensorimotor integration

It is already known that perceiving speech sounds are partly dependent on motoric information (Fadiga et al., 2002). In the second experiment, however, we found significant neural activities at the LIFG even for perceiving nonspeech sounds. Furthermore, we also found a hemodynamic modulation at the LIFG by the types of sounds. It implies that there is a sensorimotor integration to support both speech and nonspeech perception (Pulvermüller et al., 2006; Wilson et al., 2004), in which motoric information is important in top-down control of auditory perception (Scott et al., 2010). The sensorimotor integration is important for perceptual learning at the LIFG, too (Eisner et al., 2010). More specifically, the perceptual learning organizes novel sounds as perceptually meaningful elements (Westerman and Miranda, 2002; Kuhl, 2004).

The auditory-motor interface is suitable for this sensorimotor integration. By repeating in the auditory-motor interface, novel sounds may be imitated and then articulated exactly. After repeating many times, we learn how to articulate it and then perceive the sounds exactly. That is, an imitative learning eventually can shape sound perception by generating speech codes. Consistent with this notion, the LIFG is a part of the mirror neuron system used in imitative learning (Rizzolatti and Arbib, 1998). In addition, the neural circuits of imitation are very similar to the interface (Iacoboni, 2005; Iacoboni and Dapretto, 2006). In sum, in addition to the associative learning, the imitative learning is in parallel processed by the auditory-motor interface around the Sylvian fissure.

5. Right-laterality in speech processing

Last, it is worthy to mention that we found slightly right-lateralized neural activities in both experiment 1 and 3, which was not generally expected in written or visual language processing. As a plausible account of this finding, it is notable that AST (asymmetric sampling in time) hypothesis predicts this right lateralization (Poeppel, 2003). According to the AST model, unlike the classical language model, acoustic processing of speech is separately mediated by both hemispheres in different time scale: short temporal integration window for rapid frequency transitions, voice-onset time, and discriminating phonetic categories in left

hemisphere (20-40ms), and long temporal integration window for envelope peak-tracking units in right hemisphere (150-200ms).

This notion was also supported by Abrams et al. (2008). By measuring cortical-evoked potentials, Abrams et al. observed strong right hemisphere dominance for coding speech envelope in speech processing. The speech envelope was used to represent syllable pattern that was critical for normal speech perception.

Consistent with the finding, the hemispheric asymmetry was found in middle parts of posterior superior temporal sulcus (pSTS). The pSTS was suggested as a part of phonological network receiving spectro-temporally analyzed speech sounds from dorsal superior temporal gyrus (dSTG) (Hickok and Poeppel, 2007). In this sense, the right-lateralized neural activities during verbal repetition might be attributed to a contour following of the speech envelope, with larger response magnitude compared with the left homologue.

Chapter 6. Conclusion

In the present study, we explored how speech comes out of sound waves. To this end, we conducted three consecutive experiments using fMRI and fNIRS under verbal repetition paradigm, in which we revealed (1) speech has distinct neural codes modulated by high-level linguistic process specified at semantic level, irrespective of acoustic features of sounds; (2) such modulation was a result of sound learning at episodic buffer mediated by the dorsal fronto-parietal pathways from superior and middle frontal gyri (BA9, 10) to superior and inferior parietal lobules (BA7, 40) bilaterally; and (3) left inferior frontal gyrus (BA47) is pivotal for perception as well as production of meaningless/meaningful sounds in terms of brain hemodynamics specified by systolic and diastolic pulsations.

These findings are worthy in that it can provide a neural correlate of the missing linkage operating between low-level acoustic processes and high-level linguistic processes. Speech processing is in general hard to study because it is impossible to apply invasive methods used in animal studies to human brain. The present study successfully revealed a facet of online speech processing to some extent by designing novel experimental paradigms that can overcome such a limitation. At the same time, the behavioral tasks used in the study were able to recruit not only speech processing, but also memory and learning mechanism in the brain. As a result, we could describe associative learning and imitative learning,

respectively, in terms of sensorimotor integration based on auditory-motor interface.

Another contribution of this study is that we introduced verbal repetition to study speech. As shown, verbal repetition is one of simple but important tools in investigating on-line speech processing in human brain. As speech production is inextricably linked to speech perception, it is necessary to study speech processing while both ends of perception and production are active simultaneously. It is also suitable for studying auditory feedback and sensorimotor integration in speech, which are prominent features in human speech. However, lots of neurolinguistic processes in verbal repetition remain still unclear. It is thus important to reveal neural mechanism of verbal repetition, and it may shed lights on early language acquisition, too.

For future studies, we need to investigate how individual speech sounds can be organized at a phrase or sentence level. At this level, sequential processing and syntactic structure building are necessary and it will require more complex memory structures beyond the phonological memory, e.g. memory of syntactic structures or semantic (thematic) features. For such studies, we suggest that it is important to investigate a sequencing mechanism between consecutive auditory items and how a specific meaning is built from learning, which will deepen our understanding on auditory sentence processing, too. Last, it is intriguing to study which neural mechanism determines the perceptual randomness observed in the first experiment. Random switching and optimal processing of neural activities are robustly observed in the perception of ambiguous events, and as shown here, it is

deeply correlated with conceptual processing in the brain. By elucidating such phenomena, therefore, we expect to reveal a facet of brain as complex adaptive system, e.g. nonlinearity and emergence.

Chapter 7. References

Abel, K.M., Allin, M.P.G., Kucharska-Pietura, K., Andrew, C., Williams, S., David, A.S., and Phillips, M.L., 2003. Ketamine and fMRI BOLD signal: distinguishing between effects mediated by change in blood flow versus change in cognitive state. *Human Brain Mapping*, 18(2), 135-145.

Abrams, D.A., Nicol, T., Zecker, S., and Kraus, N., 2008. Right-hemisphere Auditory Cortex Is Dominant for Coding Syllable Patterns in Speech. *Journal of Neuroscience*, 28(15), 3958-3965.

Attwell, D., Buchan, A.M., Charpak, S., Lauritzen, M., MacVicar, B.A., and Newman, E.A., 2010. Glial and neuronal control of brain blood flow. *Nature*, 468, 232-243.

Bachmann, T., 2006. Microgenesis of Perception: Conceptual, Psychophysical, and Neurobiological Aspects. In H. Ögmen and B.G. Breitmeyer (Eds.), *The First Half Second*. Cambridge, Mass: MIT Press.

Baddeley, A.D. and Hitch, G.J., 1974. Working memory. In G.A. Bower (Ed.), *Recent advances in learning and motivation*, 8, 47-89, New York: Academic Press.

Baddeley, A.D., 1998. Recent developments in working memory. *Current Opinion in Neurobiology*, 8(2), 234-238.

Baddeley, A.D., 2000. The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417-423.

Baddeley, A.D., 2002. Is working memory still working? *European Psychologist*, 7, 85-87.

Baddeley, A.D., 2003a. Working memory and language: an overview. *Journal of Communication Disorders*, 36, 189-208.

Baddeley, A.D., 2003b. Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4, 829-839.

Baddeley, A.D., Gathercole, S.E., and Papagno, C., 1998. The Phonological Loop as a Language Learning Device. *Psychological Review*, 105(1), 158-173.

Bakshi, B., 1998. Multiscale PCA with application to MSPC monitoring, *American Institute of chemical Engineers Journal*, 44, 1596-1610.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., and Pike, B., 2000. Voice-selective areas in human auditory cortex. *Nature*, 403, 309-312.

Ben-Shalom, D. and Poeppel, D., 2008. Functional Anatomic Models of Language: Assembling the Pieces. *The Neuroscientist*, 14(1), 119-127.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., and Possing, E.T., 2000. Human Temporal Lobe Activation by Speech and Nonspeech Sounds. *Cerebral Cortex*, 20, 512-528.

Bloom, B.S., 1968. Learning for mastery. *Evaluation Comment*. 1(2), 1-12.

Blumenfeld, R.S. and Ranganath, C., 2006. Dorsolateral Prefrontal Cortex Promotes Long-Term Memory Formation through Its Role in Working Memory Organization. *Journal of Neuroscience*, 26(3), 916-925.

Bonte, M., Parviainen, T., Hytönen, K., and Salmelin, R., 2006. Time Course of Top-down and Bottom-up Influences on Syllable Processing in the Auditory Cortex. *Cerebral Cortex*, 16(1), 115-123.

Bozic, M., Tyler, L.K., Ives, D.T., Randall, B., and William D. Marslen-Wilson, W.D., 2010. Bihemispheric foundations for human speech comprehension. *PNAS*, 107(40), 17439-17444.

Burgess, N. and Hitch, G.J., 1999. Memory for Serial Order: A Network Model of the Phonological Loop and Its Timing. *Psychological Review*, 106(3), 551-581.

Burock, M.A., Buckner, R.L., Woldorff, M.G., Rosen, B.R., and Dale, A.M., 1998. Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *Neuroreport*, 9(16), 3735-3739.

Calder, A.J., Lawrence, A.D., and Young, A.W., 2001. Neuropsychology of Fear and Loathing. *Nature Reviews Neuroscience*, 2(5), 352-363.

Caramazza, A. and Zurif, E.B., 1976. Dissociation of algorithmic and heuristic processes in language comprehension: evidence from aphasia. *Brain and Language*, 3(4), 572-582.

Caramazza, A., 1997. How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14(1), 177-208.

Carter, A.S. and Wilson, R.H., 2001. Lexical effects on dichotic word recognition in young and elderly listeners. *Journal of American Academy of Audiology*, 12(2), 86-100.

Catani, M., Jones, D.K., and Ffytche, D.H., 2005. Perisylvian Language Networks of the Human Brain. *Annals of Neurology*, 57, 8-16.

Champod, A.S. and Petrides, M., 2007. Dissociable roles of the posterior parietal and the prefrontal cortex in manipulation and monitoring processes. *PNAS*, 104(37), 14837-13842.

Chen, X., Striano, T., and Rakoczy, H., 2004. Auditory-oral matching behavior in newborns. *Developmental Science*, 7, 42-47.

Chomsky, N., 2000. Linguistics and Brain Science. In Marantz, A., Yasushi, M., and Wayne, O. (Eds.), *Image, Language, Brain*, Cambridge, Mass: MIT Press.

Cohen, M.A., Grossberg, S., and Stork, D.G., 1988. Speech perception and production by a self-organized neural network. In Lee, Y.C. (Ed.), *Evolution, learning, and cognition*, 217-233, World Scientific Publishing Co.

Coleman, J., 1998. Cognitive reality and the phonological lexicon: a review. *Journal of Neurolinguistics*, 11, 295-320.

Cope, M. and Delpy, D.T., 1988. System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. *Medical and Biological Engineering and Computing*, 26, 289-294.

Corrigan, R., 1980. Use of Repetition to Facilitate Spontaneous Language Acquisition. *Journal of Psychological Research*, 9(3), 231-241.

Cowan, N., Elliott, E.M., Saults, J.S., Morey, C.C., Mattox, S., Hismjatullina, A., and Conway, A.R.A., 2005. On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, 51, 42-100.

Cowan, N., Fristoe, N.M., Elliott, E.M., Brunner, R.P., and Saults, J.S., 2006. Scope of attention, control of attention, and intelligence in children and adults. *Memory and Cognition*, 34, 1754-1768.

Crowder, R.G. and Morton, J., 1969. Precategorical acoustic storage (PAS). *Perception and Psychophysics*, 5, 365-373.

Cummings, A., Čeponienė, R., Koyama, A., Saygin, A. P., Townsend, J., and Dick, F., 2006. Auditory semantic networks for words and natural sounds. *Brain Research*, 1115, 92-107.

Dale, A.M. and Bruckner, R.L., 1997. Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, 5, 329-340.

Dale, A.M., 1999. Optimal Experimental Design for Event-Related fMRI. *Human Brain Mapping*, 8, 109-114.

Daunizeau, M., David, O., and Stephan, K.E., 2010. Dynamic Causal Modeling: A critical review of the biophysical and statistical foundations. *NeuroImage*, 58(2), 312-322.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., and Dehaene, S., 2005. Neural correlates of switching from auditory to speech perception. *NeuroImage*, 24, 21-33.

Denes, G., and Pizzamiglio, L. (Eds.). 1999. *Handbook of clinical and experimental neuropsychology*. Sussex, UK: Psychology Press.

Deserno, L., Sterzer, P., Wüstenberg, T., Heinz, A., and Schlagenhauf, F., 2012. Reduced Prefrontal-Parietal Effective Connectivity and Working Memory Deficits in Schizophrenia. *Journal of Neuroscience*, 32(1), 12-20.

Duncan, J., Seitz, R.J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., Newell, F.N., and Emslie, H., 2000. A neural basis for General Intelligence. *Science*, 289(5478), 457-460.

Edwards, A.D., Wyatt, J.S., Richardson, C.E., Delpy, D.T., Cope, M., and Reynolds, E.O.R., 1988. Cotside measurement of cerebral blood flow in ill newborn infants by near-infrared spectroscopy (NIRS). *Lancet*, 2, 770-771.

Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., and Scott, S.K., 2010. Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *Journal of Neuroscience*, 30(21), 7179-7186.

Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G., 2002. Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, 15(2), 399-402.

Fowler, C., 1986. An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14, 3-28.

Frackowiak, R.S.J., Friston, K.J., Frith, C.D., Dolan, R.J., and Mazziotta, J.C., 1997. *Human Brain Function*. San Diego: Academic Press.

Friedrich, F., Glenn, C., and Martin, O.S.M., 1984. Interruption of phonological coding in conduction aphasia. *Brain and Language*, 22, 266-291.

Friedrich, F., Martin, R.C., and Kemper, S., 1985. Consequences of a phonological coding deficit on sentence processing. *Cognitive Neuropsychology*, 2, 385-412.

Friston K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2, 189-210.

Friston, K.J., Harrison, L., and Penny, W., 2003. Dynamic causal modeling. *NeuroImage*, 19, 1273-1302.

Friston, K.J., Holmes, A.P., Price, C.J., Büchel, C., and Worsley, K.J., 1999. Multisubject fMRI Studies and Conjunction Analyses. *NeuroImage*, 10, 385-396.

Fukui, Y., Ajichi, Y., and Okada, E., 2003. Monte Carlo prediction of nearinfrared light propagation in realistic adult and neonatal head models. *Applied Optics*, 42, 2881-2887.

Gabrieli, J.D.E., Desmond, J.E., Demb, J.B., and Wagner, A.D., 1996. Functional magnetic resonance imaging of semantic memory processes in the frontal lobes. *Psychological Science*, 7(5), 278-283.

Gallese, V. and Lakoff, G., 2005. The Brain's Concepts: The Role of the Sensory-Motor System in Reason and Language. *Cognitive Neuropsychology*, 22, 455-479.

Genovese, C.R., Lazar, N.A., & Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, 15, 870–878.

Gerstner, W., Kreiter, A.K., Markram, H., and Herz, A.V.M., 1997. Neural codes: Firing rates and beyond. *PNAS*, 94, 12740-12741.

Ghosh, B.C.P., Calder, A.J., Peers, P.V., Lawrence, A.D., Acosta-Cabronero, J., Pereira, J.M., Hodges, J.R., and Rowe, J.B., 2012. Social cognitive deficits and their neural correlates in progressive supranuclear palsy. *Brain*, 135(7), 2089-2102.

Goldinger, S.D., 1997. Words and voices: perception and production in an episodic lexicon. In K. Johnson and J.W. Mullennix (Eds.), *Talker Variability in Speech Processing*. San Diego: Academic Press.

Gooding, P.A., Mayes, A.R., and van Eijk, R. 2000. A meta-analysis of indirect memory tests for novel material in organic amnesics. *Neuropsychologia*, 38, 666-676.

Gow, D.W. Jr., 2012. The cortical organization of lexical knowledge: a dual lexicon model of spoken language processing. *Brain and Language*, 121(3), 273-288.

Grodzinsky, Y. and Friederici, A.D., 2006. Neuroimaging of syntax and syntactic processing. *Current Opinion in Neurobiology*, 16, 240-246.

Guenther, F.H., Nieto-Castanon, A., Ghosh, S.S., Tourville, J.A., 2004. Representation of Sound Categories in Auditory Cortical Maps. *Journal of Speech, Language, and Hearing Research*, 47, 46-57.

Harrington, J., 2000. Does the Quee speak the Queen's English? *Nature*, 408, 927-928.

Henson, R.N.A., Burgess, N., and Frith, C.D., 2000. Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia*, 38, 426-440.

Herbster, A.N., Mintun, M.A., Nebes, R.D., and Becker, J.T., 1997. Regional cerebral blood flow during word and nonword reading. *Human Brain Mapping*, 5, 84–92.

Hickok, G. and Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92, 67-99.

Hickok, G. and Poeppel, D., 2007. The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393-402.

Hickok, G., Houde, J., and Rong, F., 2011. Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, 69, 407-422.

Hoekert, M., Vingerhoets, G., and Aleman, A., 2010. Results of a pilot study on the involvement of bilateral inferior frontal gyri in emotional prosody perceptionL an rTMS study. *BMC Neuroscience*, 11(93), 1-8.

Howard, D., and Nickels, L., 2005. Separating input and output phonology: Semantic, phonological, and orthographic effects in short-term memory impairment. *Cognitive Neuropsychology*, 22(1), 42-77.

Husain, F.T., Fromm, S.J., Pursley, R.H., Hosey, L.A., Braun, A.R., and Horwitz, B., 2006. Neural Bases of Categorization of Simple Speech and Nonspeech Sounds. *Human Brain Mapping*, 27, 636-651.

Hutchinson, J.B., Uncapher, M.R., and Wagner, A.D., 2009. Posterior parietal cortex and episodic retrieval: Convergent and divergent effects of attention and memory. *Learning and Memory*, 16, 343-356.

Iacoboni, M. and Dapretto, M., 2006. The mirror neuron system and the consequences of its dysfunction. *Nature Reviews*, 7, 942-951.

Iacoboni, M., 2005. Neural mechanisms of imitation. *Current Opinion in Neurobiology*, 15, 632-637.

Indefrey, P. and Levelt, W.J.M., 2004. The spatial and temporal signatures of word production components. *Cognition*, 92, 101-144.

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C., 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87, B47–57.

Jacquemot, C. and Scott, S.K., 2006. What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences*, 10(11), 480-486.

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., and Dupoux, E., 2003. Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. *Journal of Neuroscience*, 23(29), 9541-9546.

Jezzard, P. and Clare, S., 1999. Sources of Distortion in Functional MRI Data. *Human Brain Mapping*, 8, 80-85.

Johnson, K., 1997. Speech perception without speaker normalization: an exemplar model. In K. Johnson and J.W. Mullennix (Eds.), *Talker Variability in Speech Processing*. San Diego: Academic Press.

Jolliffe, I.T., 2002. *Principal component analysis*, New York: Springer.

Kemmerer, D., Castillo, J.G., Talavage, T., Patterson, S., and Wiley, C., 2008. Neuroanatomical distribution of five semantic components of verbs: Evidence from fMRI. *Brain and Language*, 107, 16-43.

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., and Gruber, O., 2009. Functional Architecture of Verbal and Tonal Working Memory: An fMRI Study. *Human Brain Mapping*, 30, 859-873.

Kuhl, P. and Rivera-Gaxiola, M., 2008. Neural substrates of language acquisition. *Annual Review of Neuroscience*, 31, 511-534.

Kuhl, P.K. and Meltzoff, A.N., 1996. Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, 100, 2425-2438.

Kuhl, P.K. and Miller, J.D., 1978. Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, 63, 905-917.

Kuhl, P.K., 2004. Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, 5, 831-843.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., and Lindblom, B., 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606-608.

Kwong, K.K., Belliveau, J.W., Chesler, D.A., Goldberg, I.E., Weisskoff, R.M., Poncelet, B.P., Kennedy, D.N., Hoppel, B.E., Cohen, M.S., and Turner, R., 1992. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *PNAS*, 89, 5675-5679.

Lammertsma, A.A., Brooks, D.J., Beaney, R.P., Turton, D.R., Kensett, M.J., Heather, J.D., Marshall, J., and Jones, T., 1984. In vivo measurement of regional cerebral haematocrit using positron emission tomography. *Journal of Cerebral Blood Flow and Metabolism*, 4, 317-322.

Lancaster, J.L., Summerln, J.L., Rainey, L., Freitas, C.S., Fox, P.T., 1997. The Talairach Daemon, a database server for Talairach Atlas Labels. *NeuroImage*, 5, S633.

Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., Nickerson, D., Mikiten, S.A., Fox, P.T., 2000. Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping*, 10, 120-131.

Lerch, D., Orglmeister, R., and Penzel, T., 2012. Automatic analysis of systolic, diastolic and mean blood pressure of continuous measurement before, during and after sleep arousals in polysomnographic overnight recordings. *Biomedizinische Technik (Berlin)*, 57, 641-644.

Levelt, W.J.M., 1989. *Speaking: From Intention to Articulation*, Cambridge, MA: MIT Press.

Levelt, W.J.M., 1992. Accessing words in speech production: Stages, processes and representations. *Cognition*, 42, 1-22.

Liberman, A.M. and Mattingly, I.G., 1985. The motor theory of speech perception revised. *Cognition*, 21(1), 1-36.

Liberman, A.M. and Whalen, D.H., 2000. On the relation of speech to language. *Trends in Cognitive Sciences*, 4(5), 187-196.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., and Studdert-Kennedy, M., 1967. Perception of the speech code. *Psychological Review*, 74, 431-461.

Liberman, A.M., Harris, K.S., Hoffman, H.S., and Griffith, B.C., 1957, The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358-368.

Lichtheim, O., 1884. On aphasia. *Brain*, 7, 443-484.

Lindquist, M.A. and Wager, T.D., 2007. Validity and Power in Hemodynamic Response Modeling: A Comparison Study and a New Approach. *Human Brain Mapping*, 28, 764-784.

Lindquist, M.A., 2008. The Statistical Analysis of fMRI Data. *Statistical Science*, 23(4), 439-464.

Liu, T.T., Franck, L.R., Wong, E.C., and Buxton, R.B., 2001. Detection power, estimation efficiency, and predictability in event-related fMRI. *NeuroImage*, 13, 759-773.

Maldjian, J.A., Laurienti, P.J., Kraft, R.A., and Burdette, J.H., 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, 19, 1233–1239 [WFU Pickatlas, version 2.3].

Malmierca, M.S. and Hackett, T.A., 2010. Structural organization of the ascending auditory pathway. In A. Rees & A.R. Palmer (Eds.), *The Oxford handbook of auditory science: The Auditory Brain*. New York: Oxford Univ. Press.

Marslen-Wilson, W., 1987. Functional parallelism in spoken word recognition. *Cognition*, 25, 71-102.

Martin, R., 2006. The neuropsychology of sentence processing: where do we stand? *Cognitive Neuropsychology*, 23, 74-95.

Massaro, D.W. and Cowan, N., 1993. Information processing models: Microscopes of the Mind. *Annual Review of Psychology*, 44, 383-425.

Massaro, D.W., 1987. *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*. Hillsdale, NJ: Lawrence Erlbaum.

McClelland, J.L., Mirman, D., and Holt, L.L., 2006. Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, 10(8), 363-369.

McNealy, K., Mazziotta, J.C., and Dapretto, M., 2006. Cracking the Language Code: Neural Mechanisms Underlying Speech Parsing. *Journal of Neuroscience*, 26(29), 7629-7639.

Newman, S.D. and Twieg, D., 2001. Differences in Auditory Processing of Words and Pseudowords: An fMRI Study. *Human Brain Mapping*, 14, 39-47.

Noesselt, T., Shah, N.J., and Jäncke, L., 2003. Top-down and bottom-up modulation of language related areas – An fMRI study. *BMC Neuroscience*, 4(13), 1-12.

Ogawa, S., Tank, D.W., Menon, R., Ellermann, J.M., Kim, S.G., Merkle, H., and Ugurbil, K., 1992. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *PNAS*, 89, 5951-5955.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1), 97-113.

Osnes, B., Hugdahl, K., Hjelmervik, H., and Specht, K., 2012. Stimulus expectancy modulates inferior frontal gyrus and premotor cortex activity in auditory perception. *Brain and Language*, 121(1), 65-69.

Papagno, C., Valentine, T., and Baddeley, A.D., 1991. Phonological short-term memory and foreign language vocabulary learning. *Journal of Memory and Language*, 30, 331-347.

Paul-Brown, D. and Soli, S.D., 1981. Interactions of lexical status and stimulus dominance effects in dichotic listening. *Journal of the Acoustical Society of America*, 69(S1), S114-S114.

Paulesu, E., Frith, C.D., and Frackowiak, R.S.J., 1993. The neural correlates of the verbal component of working memory. *Nature*, 362, 342-345.

Petkov, C.I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., and Logothetis, N.K., 2008. *Nature Neuroscience*, 11(3), 367-374.

Pinker, S., 1994. *The Language Instinct*. New York: Morrow.

Poeppel, D. and Hickok, G., 2004. Towards a new functional anatomy of language. *Cognition*, 92, 1-12.

Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, 41, 245-255.

Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., and Gabrieli, J.D.E., 1999. Functional Specialization for Semantic and Phonological Processing in the Left Inferior Prefrontal Cortex. *NeuroImage*, 10, 15-35.

Price, C.J. and Friston, K.J., 1997. Cognitive conjunction: A new approach to brain activation experiments. *NeuroImage*, 5, 261–270.

Price, C.J., Wise, R.J., Watson, J.D., Patterson, K., Howard, D., and Frackowiak, R.S., 1994. Brain activity during reading. The effects of exposure duration and task. *Brain*, 117, 1255–1269.

Pugh, K.R., Shaywitz, B.A., Shaywitz, S.E., Fulbright, R.K., Byrd, D., Skudlarski, P., Shankweiler, D.P., Katz,L., Constable, R.T., Fletcher, J., Lacadie, C., Marchione, K., and Gore, J.C., 1996. Auditory selective attention: An fMRI investigation. *NeuroImage*, 4, 159-173.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., and Shtyrov, Y, 2006. Motor cortex maps articulatory features of speech sounds. *PNAS*, 103(20), 92-101.

Ranganath, C., Johnson, M.K., and D'Esposito, M., 2003. Prefrontal activity associated with working memory and episodic long-term memory. *Neuropsychologia*, 41, 378-389.

Rauschecker, J. and Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718-724.

Rissman, J., Eliassen, J.C., and Blumstein, S.E., 2003. An event-related fMRI investigation of implicit semantic priming. *Journal of Cognitive Neuroscience*, 15(8), 1160–1175.

Rizzolatti, G. and Arbib, M.A., 1998. Language within our grasp. *Trends in Neurosciences*, 21(5), 188-194.

Rorden, C. and Brett, M., 2000. Stereotaxic display of brain lesions. *Behavioral Neurology*, 12, 191-200.

Rubin, E., 2001. Figure and Ground. In Yantis, S. (Ed.), *Visual Perception*, 225-229, Philadelphia: Psychology Press.

Sabri, M., Binder, J.R., Desai, R., Medler, D.A., Leitl, M.D., and Liebenthal, E., 2008. Attentional and linguistic interactions in speech perception. *NeuroImage*, 39(3), 1444–1456.

Schachner, A., Brady, T.F., Pepperberg, I.M., and Hauser, M.D., 2009. Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology*, 19, 831-836.

Schwarzbauer, C., Davis, M.H., Rodd, J.M., and Johnsrude, I., 2006. Interleaved silent steady state (ISSS) imaging: A new sparse imaging method applied to auditory fMRI. *NeuroImage*, 29, 774-782.

Scott, S.K., Blank, C.C., Rosen, S., and Wise, R.J.S., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123, 2400-2406.

Scott, S.K., Sauter, D., and McGettigan, C., 2010. Brain mechanisms for processing perceived emotional vocalizations in humans. In S.M. Brudzynski (Ed.), *Handbook of Mammalian Vocalization*. Oxford: Academic Press.

Smith, E.C. and Lewicki, M.S., 2006. Efficient auditory coding. *Nature*, 439, 978-982.

Soroker, N., Kasher, A., Giora, R., Batori, G., Corn, C., Gil, M., and Zaidel, E., 2005. Processing of basic speech acts following localized brain damage: A new light on the neuroanatomy of language. *Brain and Cognition*, 57, 214-217.

Stevens, K.N. and Blumstein, S.E., 1981. The search for invariant acoustic correlates of phonetic features. In P.D. Eimas and J.L. Miller (Eds.), *Perspectives on the Study of Speech*. Hillsdale: Lawrence Erlbaum.

Talairach, J. and Tournoux, P., 1988. *A co-planar stereotaxic atlas of a human brain*. Stuttgart: Theime Medical Publishers.

Techentin, C. and Voyer, D., 2011. Word frequency, familiarity, and laterality effects in a dichotic listening task. *Laterality: Asymmetries of Body, Brain and Cognition*. 16(3), 313-332.

Tourville, J.A., Reilly, K.J., and Guenther, F.H., 2008. Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, 39, 1429-1443.

Tsukiura T., Fujii, T., Fukatsu, R., Otsuki, T., Okuda, J., Umetsu, A., Suzuki, K., Tabuchi, M., Yanagawa, I., Nagasaka, T., Kawashima, R., Fukuda, H., Takahashi, S., and Yamadori, A., 2002. Neural basis of the Retrieval of People's Names: Evidence from Brain-Damaged Patients and fMRI. *Journal of Cognitive Neuroscience*, 14(6), 922-937.

Underwood, G., 1981. Lexical recognition of embedded unattended words: Some implications for reading processes. *Acta Psychologica*, 47, 267–283.

Vallar, G., Di Betta, A.M., and Silveri, M.C., 1997. The phonological short-term store-rehearsal system: Patterns of impairment and neural correlates. *Neuropsychologia*, 35(6), 795-812.

Vargha-Khadem, F., Gadian, D.G., Watkins, K.E., Connelly, A., Van Paesschen, W., and Mishkin, M., 1997. Differential Effects of Early Hippocampal Pathology on Episodic and Semantic Memory. *Science*, 277, 376-380.

Vigliocco, G., 2000. Language processing: The anatomy of meaning and syntax. *Current Biology*, 10, R78-R80.

Vouloumanos, A., Kiehl, K.A., Werker, J.F., and Liddle, P.F., 2001. Detection of Sounds in the Auditory Stream: Event-Related fMRI Evidence for Differential Activation to Speech and Nonspeech. *Journal of Cognitive Neuroscience*, 13(7), 994-1005.

Wagner, A.D., Shannon, B.J., Kahn, I., and Buckner, R.L., 2005. Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, 9(9), 445-453.

Wallesch, C.W. and Kertesz, A., 1993. Clinical symptoms and syndromes of aphasia. In G. Blanken, J. Dittmann, H. Grimm, J.C. Marshall, and C.-W. Wallesch (Eds.), *Linguistic disorders and pathologies*, 98-119, Berlin: de Gruyter.

Wang, L., Liu, X., Guise, K.G., Knight, R.T., Ghajar, J., and Fan, J., 2009. Effective Connectivity of the Fronto-parietal Network during Attentional Control. *Journal of Cognitive Neuroscience*, 22(3), 543-553.

Warren, J.E., Sauter, D.A., Eisner, F., Wiland, J., Alexander Dresner, M., Wise, R.J.S., Rosen, S., and Scott, S.K., 2006. Positive emotions preferentially engage an

auditory-motor “mirror” system. *Journal of Neuroscience*, 26(50), 13067-13075.

Weekes, N.Y., Capetillo-Cunliffe, L., Rayman, J., Iacoboni, M., and Zaidel, E., 1999. Individual differences in the hemispheric specialization of dual route variables. *Brain and Language*, 67(2), 110-133.

Westerman, G. and Miranda, E.R., 2002. Modelling the development of mirror neurons for auditory-motor integration. *Journal of New Music Research*. 31(4), 367-375.

Wilson, S. and Iacoboni, M., 2006. Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *NeuroImage*, 33(1), 316-325.

Wilson, S., Saygin, A., Sereno, M., and Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7(7), 701-702.

Yates, J. and Thul, N., 1979. Perceiving surprising words in an unattended auditory channel. *Quarterly Journal of Experimental Psychology*, 31, 281–286.

Yoncheva, Y.N., Zevin, J.D., Maurer, U., and McCandliss, B.D., 2010. Auditory selective attention to speech modulates activity in the visual word form area. *Cerebral Cortex*, 20(3), 622-632.

Yonsei Korean Dictionary, 1998. *Yonsei Korean Corpus 1-9*, Institute of Language and Information Studies, Yonsei University, Korea.

Yoo, S., Chung, J-Y., Jeon, H-A., Lee, K-M., Kim, Y-B., and Cho, Z-H., 2012. Dual routes for verbal repetition: Articulation-based and acoustic-phonetic codes

for pseudoword and word repetition, respectively. *Brain and Language*, 122(1), 1-10.

Zaidel, E., Clarke, J.M., and Suyenobu, B., 1990. Hemispheric independence: A paradigm case for cognitive neuroscience. In A.B. Scheibel and A.F. Wechsler (Eds.), *Neurobiology of Higher Cognitive Function*, 297-355, NY: Guilford Press.

Zhang, D., Zhang, X., Sun, X., Li, Z., Wang, Z., He, S., and Hu, X., 2004. Cross-Model Temporal Order Memory for Auditory Digits and Visual Locations: An fMRI Study. *Human Brain Mapping*, 22, 280-289.

Zhang, Y. and Wang, Y., 2007. Neural plasticity in speech acquisition and learning. *Bilingualism: Language and Cognition*, 10(2), 147-160.

Ziefle, M. 1998. Effects of display resolution on visual performance. *Human Factors*, 40(4), 555-568.

Chapter 8. Appendix

1. Behavioral evaluation of repeating ambiguous speech sounds

During verbal repetition of ambiguous speech sounds, the perceptual processes would resemble those for bi-stable perceptions in visual domain (Rubin, 2001). In the vase/profile illusion image, one feature is not more salient than the other, i.e. a vase and a profile of human faces are inextricably intermingled with each other by having a common border. As a result, the perceived image is spontaneously determined by attending to a specific feature and it leads to figure-ground reversals. That is, the perceived image randomly fluctuates between two possible candidates, a vase or a profile. It is impossible to perceive both meaningful images at the same time.

Similarly, if one phonetic feature is spontaneously attended to, it pops out as the figure and is uniquely perceived while the unattended one is represented as the ground and hardly perceived during auditory perception. What we perceive from the ambiguous sounds is randomly determined by spontaneous attention whenever one hears the sounds. However, if one perceives two distinct sounds that are unequal in saliency, one sound may be favored over the other. In this context, to examine the behavioral properties of the ambiguous sounds in some

detail, we conducted an experiment where two sounds were biased in terms of the lexical representation.

Six native Korean speakers (2 females and 4 males) aged 18–25 years old (mean 21.8 years) participated in the preliminary experiments. Written and informed consent was obtained from all subjects before the experiment. All subjects were strongly right-handed as assessed by the Edinburgh handedness inventory (Oldfield, 1971) and had normal auditory ability and no neurological or medical disorders. We asked the participants to repeat 84 pairs of words that differed in occurrence frequency in a Korean corpus (Yonsei Korean Dictionary, 1998). The pairs were mixed by the same procedure as described in the Method section. In all trials, the subjects should repeat immediately what they heard and report their responses in 2AFC (two-alternative forced choice; Res1 and 2) manner.

The results were summarized in Figure S1. In all subjects, the responses were biased toward one of two candidate sounds, i.e. Res1 (see Figure S1, left), where the abscissa indicates the subjects while the ordinate indicates the ratio of each response. It might be due to word frequency (or familiarity), lexical status, and stimulus dominance effects (Paul-Brown and Soli, 1981; Carter and Wilson, 2001; Techentin and Voyer, 2011). The subjects were likely to attend to the temporal profile of more familiar words in the ambiguous stimuli, leading to make their phonetic features salient and activating the perception in action via selective auditory attention (Pugh et al., 1996; Yoncheva et al., 2010). This bias is in part attributable for the homogeneity of the subjects with similar linguistic and educational backgrounds.

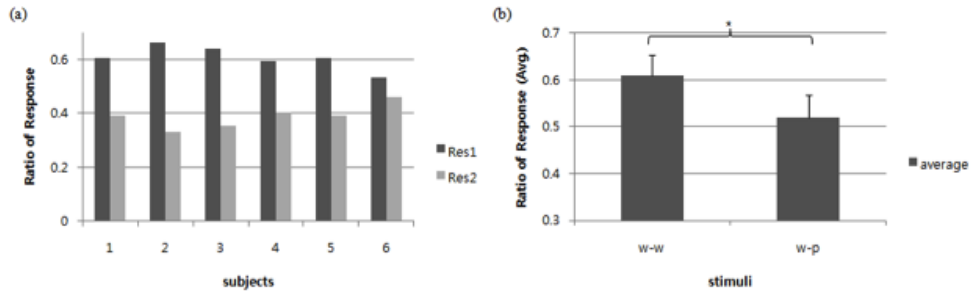


Figure S1. Responses from six subjects

In 2AFC manner, the subjects reported what they heard. (a) In all subjects, the ratio of response was biased toward one of two candidates, i.e. Res1. The abscissa indicates the subjects while the ordinate indicates the ratio of each response. (b) For comparisons, we depicted the averaged ratio of response obtained from both the preliminary experiments (word-word mixtures, w-w) and the present study (word-pseudoword mixtures, w-p) in the same graph (w-w: 0.61±0.045, w-p: 0.52±0.048). The difference between two experiments was statistically significant at the 99 % confidence level (*t*-test, $p = 0.001$).

The averaged ratio of response clearly showed that the subjects preferred one sound to the other (see Figure S1, right). For comparisons, we depicted the averaged ratio of response obtained from both the preliminary experiments (word-word mixtures, w-w) and the present study (word-pseudoword mixtures, w-p) in the same graph (w-w: 0.61±0.045, w-p: 0.52±0.048). The difference between two experiments was statistically significant at the 99 % confidence level (*t*-test, $p = 0.001$). Unlike the word-pseudoword mixtures used in the present study, the response to the word-word mixtures was statistically unbalanced.

As the data from the preliminary experiments with only six subjects might not be enough for the behavioral data analysis, the bootstrap sampling was additionally used to estimate the standard error of the median by repeatedly drawing bootstrap samples from the data. We computed a sample of 1000 bootstrapped means of random samples taken from the ratio of response, and

plotted an estimate of the density of these bootstrapped means (see Figure S2). The result shows that the difference between two types of stimuli, i.e. word-pseudoword mixtures (w-p, solid line) and word-word mixtures (w-w, dotted line), is consistently observed in the bootstrap sampling data, too.

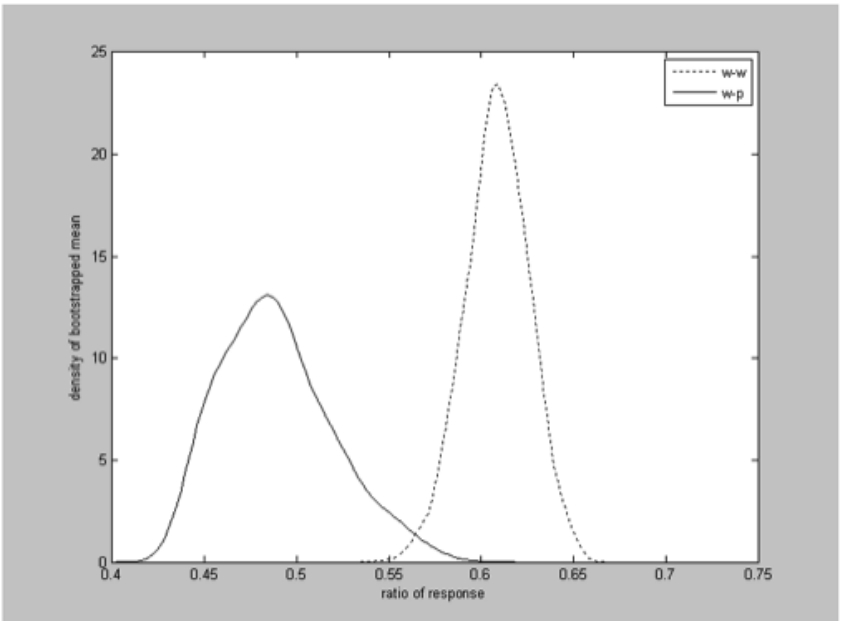


Figure S2. Bootstrapped means of random samples

Bootstrapped means of random samples taken from the ratio of response, e.g. Res1 (Total 1000 samples): word-pseudoword mixtures (w-p, solid line), and word-word mixtures (w-w, dotted line).

2. Reliability of subjects' responses

As described, a subject's response may be changed even in the same stimulus. However, it is not a problem but rather an intended condition in the present study. What we aim to do here is that the subjects are forced to recognize one distinct sound after listening to the ambiguous sound originated from the

phonologically-similar pairs of word and pseudoword. This does not mean that they always hear the same sound from the ambiguous stimulus. If they really do, it implies that the stimuli were not ambiguous enough to evoke phonetic features of both word and pseudoword simultaneously. It is again analogous to the Rubin vase/profile illusion or Necker cube that demonstrates the ambiguity in visual perception. The perceived object is not stable on one-featured space but automatically changed from figure to ground, and vice versa (figure-ground reversals). In this sense, the subjects' responses are intrinsically bi-stable and thus spontaneously changed by instant perception. We observed it in the result that the subjects' responses were actually random as depicted in Figure 5. What is important here is not whether they always perceived the same sound from one ambiguous sound but whether word- and pseudoword-perceived trials were differentially processed in the brain even though they consisted of the same acoustic sounds.

3. Rapid functional MRI for repetition task

In general, it is known that the design efficiency of event-related functional MRI study is significantly reduced as the ISI (inter-stimulus interval) decreases. The problem becomes more severe in case of rapid event-related design. Nevertheless, we adopted the rapid event-related fMRI design with ISI of 2 sec in this study because we aimed at investigating speech perception and

production with both ends active simultaneously, i.e. *in vivo* language. Since the perception and production are intermingled with each other, we used very simplified task paradigm (repetition task) requiring rapid event-related design to minimize the problem (Refer to the introduction part for the advantages of repetition tasks as a tool to study *in vivo* language).

It is likely to separate BOLD (blood oxygenation level-dependent) responses evoked by two adjacent or temporally very close stimuli if we de-convolve two conditions relevantly with separated HRFs (hemodynamic response functions) and randomize the stimuli by using jitter or different stimulus types. The co-linearity due to slow latency of BOLD response can be sufficiently removed by randomizing the onset of HRFs. In practice, two seconds of fixed ISI was successfully used to detect differential activation by speech and nonspeech in event-related fMRI design (Vouloumanos et al., 2001), and in case of visual task, even very short ISI of 500 msec in a fixed ISI scheme was enough to discriminate stimuli in left and right hemi-fields only if the stimuli were randomly presented (Burock et al., 1998).

Nonetheless, we used a fixed ISI design, instead of randomized ISI, to prevent high-level linguistic functions or any other mental processes from being explicitly involved in the task. If we randomized the ISI, the subjects should wait for some duration of randomized time after listening to the stimulus. In the additional time, there may be unintended cognitive processes, e.g. subvocal rehearsal, short-term memory management, retrievals of phonologically- and semantically-correlated words because two adjacent conditions (perception and

production) are correlated with each other unlike the typical event-related designs. Consequently, all trials are not equally processed because of these confounding factors. To avoid the problem, we fixed the ISI duration so that we could keep the homogeneity of all trials. Interestingly, the production phase, followed by the perception phase, has no external stimuli, and the type of stimuli in the production phase is *randomly* determined by the subjects' perception in the preceding phase as described in the above. In this sense, our experiment design is similar to those indicated in the above, in that it has a fixed ISI duration with randomized presentation.

However, it is not certain that the design efficiency in rapid event-related fMRI adopting repetition task is really comparable to the studies cited in the above. No published data is available now. Furthermore, we introduced newly designed experimental paradigm and stimuli with different modalities as well. For this reason, we treated the perception and production as one single event in the present study to avoid misreading of the experimental results. The difference between word- and pseudoword-perceived repetitions was not affected or compromised by doing so. The only penalty is that we cannot investigate the contrasts between perception and production during the tasks.

4. Stimuli List

1. Experiment 1: Word-Pseudoword pairs

Trials	Run 1		Run 2		Run 3		Run 4	
	WRD	PWD	WRD	PWD	WRD	PWD	WRD	PWD
1	머리	모리	나라	나러	작업	적업	노력	너력
2	사람	사럼	하루	호루	건물	건몰	고통	고텅
3	얼굴	올굴	사이	소이	노동	너동	방안	병안
4	다음	다엄	오늘	어늘	누나	누너	입술	입설
5	아들	오들	아침	오침	외국	외걱	시골	시걸
6	동안	덩안	운동	운덩	약속	약석	통일	텅일
7	순간	순건	표정	표종	동물	덩물	공간	겅간
8	국가	걱가	신문	신먼	도로	도러	처지	추지
9	문학	먼학	결국	결걱	지도	지더	전부	전버
10	행동	행덩	노래	너래	방문	방먼	시험	시홈
11	바다	바더	목적	목족	공사	겅사	구멍	구몽
12	녀석	녀속	도시	더시	행복	행벅	최고	최거
13	소설	소솔	감정	감종	목숨	먹숨	성적	성족
14	성격	송격	활동	할동	전기	존기	근대	긴대
15	생명	생멍	고향	거향	전국	전걱	성질	송질
16	물건	물곤	조선	저선	결론	결런	시설	시솔
17	공장	겅장	과거	과고	조각	저각	동기	덩기
18	부모	부머	어둠	오둠	침묵	침먹	최초	최처
19	손님	선님	사업	사옵	사물	사멀	문명	문멍
20	여성	여송	언어	언오	음성	음송	공포	공퍼
21	음악	음억	시선	시손	지붕	지벙	증거	증고

WRD: words, PWD: pseudowords; totally 84 pairs

2. Experiment 2: Words and Pseudowords only

Trials	Words	Pseudowords
1	고등학교	파하가자
2	공산주의	하아사타
3	국민학교	라사파자
4	국회의원	타카자라
5	마찬가지	차나자카
6	머리카락	파하카자
7	민주주의	파자가나
8	부끄러움	파사타다
9	사회주의	아자가바

10	시어머니	가사다나
11	아나운서	다가자라
12	아름다움	아차사나
13	아주머니	사타파라
14	어린아이	다아자바
15	여러가지	다하나자
16	오랫동안	자가하카
17	우리나라	카가파하
18	울음소리	하파카아
19	자본주의	다바나가
20	제국주의	하마사차

3. Experiment 3: Pseudowords & Reading passages

1. Verbal materials

트롱바, 쿠차이, 래이션, 뚜시타, 도쿠사 밍나주, 포바스, 하카스, 말드이, 니주미

2. Reading materials

Passage 1: Rapunzel story

옛날에 한 부부가 살았다. 뚜시타 그들은 아이가 없었다. 뚜시타 그러던 어느 날 아내가 임신을 했다. 뚜시타 아내는 요정의 들상추를 먹고 싶었다. 트롱바 아내를 몹시 사랑한 남편은 들상추를 훔쳐 왔다. 트롱바 그러다 남편은 요정에게 들키고 말았다. 뚜시타 남편은 요정에게 용서를 구했다. 트롱바 요정은 용서해 주는 대신 아이를 요구했다. 트롱바 겁이 난 남편은 엉겁결에 약속하고 말았다. 뚜시타 아이를 낳자 요정은 아이를 데려갔다. 말드이 요정은 아이를 라푼젤로 불렀다. 말드이 라푼젤은 세상에서 가장 예쁜 아이로 자랐다. 트롱바 요정은 라푼젤을 입구가 없는 탑에 가뒀다. 말드이 라푼젤의 머리카락은 아주 길고 아름다웠다. 말드이 요정은 라푼젤의 긴 머리카락을 이용하여 탑에 올랐다. 래이션 어느 날 한 왕자가 라푼젤의 노래를 들었다. 래이션 왕자는 라푼젤을 만나고 싶었지만 탑에 오를 수 없었다. 포바스 고민하던 왕자는 요정이 탑에 오르는 모습을 보았다. 포바스 결국 왕자는 라푼젤을 만나서 사랑하게 되었다. 포바스 그러나 요정에게

들킨 라푼젤은 황야로 내쫓겼다. 래이션 이를 모르고 탑에 오른 왕자는 절망하여 탑에서 뛰어내렸다. 래이션 상처로 실명한 왕자는 라푼젤을 찾아 다녔다. 포바스 하루는 라푼젤이 쌍둥이와 함께 살고 있는 황야에 이르렀다. 래이션 왕자를 알아본 라푼젤이 그의 품에 안겼다. 포바스 그때 라푼젤의 눈물이 왕자의 눈을 적셨다. 말드이 그러자 왕자는 다시 볼 수 있게 되었다

Passage 2: Short diary

안녕하세요. 내 이름은 미포틱입니다. 나에게는 정말 사랑하는 쿠차이가 있습니다. 쿠차이는 우리나라에서 아빠를 뜻하는 말입니다. 우리 쿠차이는 올해 40 세이시고 지난 13 년 동안 밍나주로 일해 오셨습니다. 밍나주는 공무원을 말하는데, 우리 쿠차이는 시청에서 도시계획 업무를 하십니다. 보통 밍나주의 월급은 많지 않지만 누구보다 열심히 일해 오신 것을 잘 알기에 우리 가족은 밍나주이신 쿠차이를 존경하고 사랑합니다. 보통 휴일에는 온 가족이 함께 도쿠사에 가는데 도쿠사는 일종의 놀이공원입니다. 오늘은 밍나주의 날을 맞이해서 온 가족이 집에서 30 분 거리에 있는 도쿠사로 놀러 갔습니다. 도쿠사에서 내가 제일 좋아하는 것은 니주미라고 하는 분장쇼입니다. 니주미에는 보통 귀여운 동물탈을 쓴 사람들이 나오는데 오늘은 특이하게 하카스를 닮은 사람이 있었습니다. 하카스는 요즘 유행하는 만화에 나오는 고릴라인데 너무 못생겨서 나는 싫어합니다. 그렇지만 대다수의 아이들에게 하카스의 인기는 정말 좋습니다. 아마도 이런 인기 덕택에 니주미에 등장하는 것 같습니다. 오늘 니주미에서도 다른 동물들보다 하카스에 대한 관심이 제일 많았습니다. 니주미를 보고 난 후에 나는 앞으로 하카스가 좋아질 것 같은 생각이 들었습니다. 다같이 점심을 먹고 도쿠사에 있는 놀이기구를 타면서 신나게 놀다 보니 어느덧 해가 지기 시작했습니다. 이처럼 온 가족이 함께 하는 나들이는 언제나 재미있습니다.

국문 초록

따라말하기는 음성 인식 및 발화의 양단을 동시에 포함하고 있으며, 특히 단어 학습에서 언어 습득의 기본 도구가 되는 간단하고 자연스러운 과제 중의 하나이다. 본 연구에서 우리는 (1) 인간의 뇌에서 음성언어코드가 어떻게 표상되는지, (2) 첫 번째 실험 결과를 지지하는, 다양한 청각자극의 수동적 듣기 및 따라말하기 과제를 수행하는 동안 관찰되는 뇌혈류역학, 그리고 (3) 음성언어코드를 만들기 위해서 새로운 소리를 특정 의미와 연합시키는데 동원되는 신경회로를 연구하는 것을 목표로 했다.

첫 번째 실험에서, 우리는 소리에 포함된 모호한 모음의 해석에 따라 단어나 비단어로 지각되는 새로운 소리자극을 사용하였다. 우리는 사건관련 기능적자기공명영상을 이용하여 좌우 뇌의 실비우스틈새와 상측두고랑에서 청각-조음 인터페이스를 발견하였다. 더 중요하게, 단어로 지각되는 시행과 비단어로 지각되는 시행의 대조를 통해 우리는 좌측 후중측두이랑에서 단어로 지각되는 소리의 따라말하기에 특징적인 신경활동을 발견하고 좌측 하전두이랑에서는 비단어로 지각할 때에 특징적인 신경활동을 발견하였다. 이러한 발견은 두 가지 독립된 음성코드 – 비단어를 위한 조음 기반의 코드와 단어를위한 음향-음성 코드 – 에 의해 지각된 소리가 의미를 가졌는지의 여부에 따라 따라말하기에서 다르게 사용되는 것을 암시한다.

두 번째 실험에서, 우리는 첫 번째 실험의 결과를 근적외선분광법에 의해 측정된 뇌혈류역학과 관련하여 다시 검증해 보았다. 우리는 피험자가 다양한 소리들, 즉 자연소리, 동물소리, 인간의 감정적 소리, 비단어, 단어를 듣고 음성언어들(비단어와 단어)만을 따라말하는 동안 좌우 하전두이랑에서 헤모글로빈의 변화를 관찰하였다. 관찰 결과, 우리는 좌측 하전두이랑에서 산화헤모글로빈의 변화가 단어와 비단어 모두에 대해 증가하지만, 우측 하전두이랑에서는 감소하는 것을 발견하였다. 더군다나, 수동적 듣기임에도 불구하고 좌측 하전두이랑에서 소리의 유형에 따라 뇌혈류역학적 변화가 달라지는 것을 관찰하였다. 단어와 비단어 따라말하기 조건만을 비교해 본 결과, 전체 헤모글로빈 변화에서 산화헤모글로빈의 증가 비율이 단어보다 비단어 경우에 유의미하게 높았는데, 이는 좌측 하전두이랑에서 조음기반 코드가 단어가 아닌, 비단어를 위해 두드러지게 사용됨을 암시한다.

세 번째 실험에서, 우리는 더 나아가서 소리가 어떻게 의미를 가지게 되는지 즉 어떤 신경기제가 그 소리를 의미화하는 학습 과정을 지원하는지를 조사하였다. 우리는 기능성자기공명영상과 결합된 간단한 연합학습 패러다임을 설계하였다. 연합학습을 위해서, 몇 개의 새로운 소리들이 그 의미와 함께 간단한 이야기 를 통해 제시된 반면 (학습조건), 다른 몇몇 소리들은 동일한 이야기 속에서 의미 설명 없이 제시되었다 (비학습조건). 학습 단계 전후의 따라말하기를 비교해 본 결과, 학습되지 않은 소리들의 따라말하기는 좌우 상전두이랑과 중전두이랑에서 특징적인 신경활동을 보여준 반면, 학습된 소리들은 좌우 상두정엽과 하두정엽에서

특징적인 신경활동을 나타내었다. 동적인과모형을 사용하여 연결성 분석을 한 결과, 등쪽 전두-두정 네트워크가 일화버퍼로서 새로운 소리의 연합학습을 위해 사용됨을 알 수 있었다.

모든 결과를 종합해 볼 때, 우리는 등쪽 전두-두정 네트워크가 새로운 소리를 음성언어소리, 즉 의미있는 소리로 변환하는 연합학습에 동원되는 것을 발견하였다. 어떤 소리가 연합학습에 의해 의미를 지니게 되면, 좌측 중측두이랑에서 음향-음성 코드로 표상되어 사용되는 반면, 의미없는 소리는 좌측 하전두이랑에서 일시적으로 조음기반 코드를 생성하여 유지된다. 이런 연구 결과는 좌측 하전두이랑에서 음성언어 사용 시의 뇌혈류역학적 변화에 의해 확인되었는데, 이는 음성언어의 지각이 부분적으로는 음성발화를 위한 음성언어코드의 생성에 의존한다는 것을 암시한다.

주요어: 따라말하기, 단어 학습, 기능적자기공명영상, 기능적근적외선분광법, 음성언어코드, 동적인과모형

학번: 2009-30811