Cortical mechanisms underlying the subjective perception of spectrally degraded speech

by

Shota Murai

Doshisha University
Graduate School of Life and Medical Sciences

November 2021

Doshisha University
Graduate School of Life and Medical Sciences

Abstract

Cortical mechanisms underlying the subjective perception of spectrally degraded speech

by Shota Murai

Speech perception is one of the most important aspects of interpersonal communication in everyday life. However, speech sounds can be degraded by noisy environmental sounds and sensorineural hearing loss. Therefore, understanding the central nervous system's role in the perceptual process of speech sounds under acoustically degraded listening conditions would be vital in enhancing the auditory experience and improving the perceptual performance of speech sounds. The purpose of the current thesis is to characterize neural substrates that underlie the variable perceptual processes of degraded speech sounds within individuals, focusing on noise-vocoded speech sounds (NVSS), simulating sensory inputs of cochlear implant recipients. In the first study, functional magnetic resonance imaging (fMRI) measurements investigated the relationship between the fluctuation of degraded speech comprehension within individuals and the cerebral activity. Our results indicate that higher comprehension of NVSS sentences was associated with greater activation in the right superior temporal cortex.

Additionally, activity in the left inferior frontal gyrus (Broca's area) was increased when a listener recognized words in a sentence they did not fully comprehend. In addition, results of laterality analysis demonstrated that recognition of words in an NVSS sentence led to less lateralized responses in the temporal cortex, though a left-lateralization was observed when no words were recognized. In the second study, the nature of perceptual learning processes of acoustically degraded speech was considered.

The lengthy learning period underlies the rehabilitation of patients with hearing aids or cochlear implants. To reveal the neural processes occurring during and after long-term perceptual learning, perceptual training of NVSS with fMRI measurements was conducted for seven days and for the postsession (approximately one year later). Behavioral results demonstrated that participants improved their performance across seven experimental days; this improvement was maintained in the postsession. Furthermore, representational similarity analysis showed that the neural activation

patterns to NVSS relative to clear speech in the left posterior superior temporal sulcus (pSTS) on Day 7 got considerably different from Day 1, accompanying neural changes in frontal cortices. In addition, the distinct activation patterns to NVSS in the pSTS were also observed in the postsession. These results suggest that neural changes in the temporal regions improved and maintained degraded speech perception. These behavioral improvements and neural changes induced by the perceptual learning of degraded speech will provide insights into cortical mechanisms underlying adaptive processes in difficult listening situations and long-term rehabilitation of auditory disorders. In short, variation in degraded speech perception within individuals is underpinned by the perceptual processing in the frontotemporal regions, as reflected by cortical lateralization and changes in neural representation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

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

First, I would particularly like to thank Professor Kohta I. Kobayasi for providing me this invaluable opportunity to study as a doctoral student, for his constructive comments, and for encouraging me with his grateful support throughout years. I would also like to express the deepest appreciation to Professor Hiroshi Riquimaroux for insightful advices, ceaseless great support and warm encouragement. I wish to thank Professor Shizuko Hiryu for her grateful supports and shrewd advices to my research. I would like to express my gratitude to them for providing such a great research opportunity for my thesis work.

I would like to express my sincere gratitude to Dr. Ryosuke Tachibana, Dr. Jan Auracher, Dr. Emyo Fujioka, Dr. Yasufumi Yamada, Dr. Takafumi Furuyama, Dr. Hidetaka Yashiro, Dr. Miwa Sumiya, Dr. Kazuma Hase, Dr. Sachi Itagaki, Dr. Yuta Tamai, Mr. Yuma Osako and Mr. Yuki Ito for their shrewd advices and warm encouragement, and Mr. Yuki Nakayama, Ms. Marina Takabayashi, Mr. Genki Asaka, Mr. Takuma Kitayama, Ms. Ae Na Yang, Mr. Takuma Matsumura, Ms. Momoka Nishimura, Mr. Makoto Matsumoto, Mr. Jun Shimpaku, Ms. Momoko Hishitani, Mr. Hidekazu Nagamura, Mr. Hiroshi Ohnishi, Ms. Ayuna Tamura, Mr. Tamaki Noguchi, and Ms. Tomomi Watanabe, and all of the members of the Sensory and Cognitive Neural System Laboratory and the Neuroethology and Bioengineering Laboratory in Doshisha University for their support.

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

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

Shota Murai

1. General introduction

Speech perception is one of the most important aspects of interpersonal communication and is crucial for the quality of life. Sensory inputs underlie speech perception process complex auditory information like amplitude envelope and temporal fine structures. Speech sounds can be, however, degraded by noisy environmental sounds and sensorineural hearing loss. Therefore, understanding the central nervous system's role in the perceptual process of speech sounds under acoustically degraded listening conditions would be vital in enhancing the auditory experience and improving speech sounds' perceptual performance. The current thesis reports neuroimaging research on the perception of degraded speech sounds, focusing mainly on noise-vocoded speech sounds (NVSS), simulating sensory inputs of cochlear implant recipients. This introductory chapter begins with a review of the prior literature on degraded speech perception, shedding light on acoustic and neural characteristics of the process of speech perception. The hypothesis and objective of each of the following chapters are then discussed.

1.1. Perception of degraded speech

The human speech perceptual system exhibits robustness under acoustically degraded listening situations. Psychoacoustical studies investigated the perceptual process of various types of degradation of speech sounds, such as sine-wave speech (Remez, Rubin, Pisoni, & Carrell, 1981), time-compressed speech (Fairbanks & Kodman, 1957), locally time reversed speech (Saberi & Perrott, 1999), mosaic speech (Nakajima, Matsuda, Ueda, & Remijn, 2018), and NVSS. NVSS allows us to simulate auditory signal degradation of

cochlear implants and investigate the role of amplitude envelope and spectral detail in speech perception.

Noise-vocoded speech sounds

Shannon and colleagues developed a vocoding technique to produce NVSS that speech sounds were spectrally degraded in a controlled manner (Robert V. Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). NVSS is synthesized speech sounds where multiple noise bands replace the speech signal while the amplitude envelope is primarily preserved, resulting in greatly reduced spectral detail within each band. First, the amplitude envelopes were extracted through several frequency bands from the original (nonvocoded) speech signal. Second, the envelopes were extracted by half-wave rectification and low-pass filtering (cutoff frequencies: 16, 50, 160, or 500 Hz). The effect of temporal detail in the amplitude envelope was evaluated by manipulating the cutoff frequency used for the low-pass filters. Third, the envelope of each frequency band was used to modulate the white noise of the same frequency band. NVSS were then synthesized by summing these modulated noises. The fundamental frequency was not present, and formant peaks were not identified in NVSS.

Speech recognition performance of vowels, consonants, and sentences was measured. As a result, although participants listened to speech sound with a reduction in spectral content under the conditions of four-band modulated noise, the preserved temporal cues were sufficient to produce approximately 90% correct response rate from the identification of words. The results also indicated that the performance on all speech measurements increased with the number of frequency bands and the cutoff frequency of amplitude envelopes. However, when the cutoff frequency was the lowest (i.e., 16 Hz), a

significant reduction in performance was observed for consonants and sentences, suggesting slow amplitude envelope information is a critical cue for speech perception under spectrally degraded conditions. The results thus revealed that minimal spectral information is required for speech perception when temporal cues are available in a few spectral bands. This also suggested alternative signal-processing strategies for auditory prostheses.

The artificial degradation of speech signals allows us to control the intelligibility of speech sounds. For example, if the number of frequency bands is smaller, the spectral information of speech sounds is less detailed and the sounds are less intelligible. The technique has been widely applied to behavioral and neuroimaging studies (e.g., Davis & Johnsrude, 2003; Scott, Blank, Rosen, & Wise, 2000; Xu, Tsai, & Pfingst, 2002).

Perceptual cues in degraded speech

The robustness against limited spectral cues with preserved temporal cues is observed in multiple languages. The Fu et al. study extended the findings in English speech to a tonal language such as Mandarin Chinese speech (Fu, Zeng, Shannon, & Soli, 1998). Tonal languages, such as Thai, Cantonese, Vietnamese, and Mandarin Chinese, have temporal fundamental frequency variations (F0) that express lexical meaning within isolated syllables. Native Chinese-speaking listeners were asked to identify Chinese vowels, consonants, tones, and sentences of NVSS comprising one, two, three, and four noise bands. The results showed that the percentage of correct recognition of vowels, consonants, and sentences increased with noise bands.

Furthermore, approximately 80% of Chinese tones were correctly recognized, independent of the number of frequency bands. Riquimaroux (2006) demonstrated the

perception of Japanese speech sounds using NVSS. The intelligibility increased as a function of the number of bands, and four-band NVSS are mostly intelligible (about 80% correct recognition ratio in a sentence), consistent with English and Chinese speech observation. The accent of Japanese words was also perceivable even though NVSS does not contain F0. Accent (change in pitch) of Japanese words is usually characterized by temporal patterns of F0 within a word and indicates lexical differences. The author provided a potential explanation for the accent perception in NVSS that the central nervous system would translate changes in auditory input of the amplitude envelope into changes in perception of changes in F0. It is important to note that accents could be carried by the amplitude envelope only when speech signals were replaced by noise bands and did not contain F0 information. As long as F0 information is available, the amplitude envelope cannot carry accents; however, the amplitude envelope would support accent's perception under a spectrally degraded condition.

Ueda and Nakajima (2017) revealed the number and frequency ranges of frequency bands for efficient speech perception across eight different languages/dialects (American English, British English, Cantonese, French, German, Japanese, Mandarin, and Spanish). Using the principal component analysis, the results of the four-factor analysis indicated four common frequency bands, including 50–540, 540–1,700, 1,700–3,300, and 3,300–7000 Hz. The results of the three-factor analysis revealed three factors: the frequency ranges of 50–540 and 1,700–3,300 Hz, the range of 540–1,700 Hz, and the range above 3,300 Hz. These results indicated universally appeared acoustic components of speech sounds (Ueda & Nakajima, 2017). Behavioral experiments were conducted with four-band NVSS to investigate the efficacy of the number and frequency ranges of frequency bands obtained from the factor analyses (Ueda, Araki, & Nakajima, 2018). The

mean mora percent correct score across participants exceeded more than 90% without training. In addition, combining and exchanging amplitude envelope patterns of the frequency bands in four-band NVSS led to a declining score. These behavioral results further revealed robustness and frequency specificity of amplitude envelope contributions to NVSS perception.

Segmental and suprasegmental information in NVSS would thus largely be carried by amplitude envelope when the spectral resolution was limited. Xu and colleagues (2005) demonstrated the effect of temporal and spectral resolution on phoneme recognition of NVSS. Vowel and consonant recognition scores were collected from native-English-speaking listeners while systematically varying the amount of spectral and temporal information of NVSS (Xu, Thompson, & Pfingst, 2005). Concerning amplitude envelope, the results revealed that the vowel and consonant recognition score improved up to 4- and 16-Hz cutoff frequency filters. Eight and six spectral bands were needed to achieve plateau recognition performance. These results suggested the contribution of amplitude envelope and spectral details to phonemic recognition and some tradeoff between these two acoustic cues. Tachibana and colleagues (2013) extended the interaction between temporal and spectral resolution to multiple levels of the hierarchy of the linguistic structure, such as syllable, word, and sentence. Syllables, words, and sentences of NVSS were presented to participants, with a systematic manipulation of amplitude envelope, altered by low-pass filters (4-, 8-, or 16-Hz cutoff frequency) and spectral information, controlled by the number of spectral bands (4, 8, or 16 bands) (Tachibana, Sasaki, & Riquimaroux, 2013). The benefit from amplitude envelope and spectral information was evaluated by comparing the accuracy of recognition of NVSS. The results demonstrated that the NVSS recognition benefited more from amplitude

envelope in a higher level of a hierarchy of linguistic structure. This suggested that amplitude envelope in NVSS reflects suprasegmental information conveying prosodic properties of the speech signal, such as coarticulation, accent, and intonation, used in lexical and syntactic processing.

Taken together, under the spectrally degraded conditions, the process of speech perception is underpinned by the interactions between amplitude envelope and spectral information to achieve successful comprehension.

Perceptual learning of degraded speech

How can the human speech perceptual system capture the temporal and spectral cues from speech sounds even under acoustically degraded conditions different from daily life? Davis and colleagues (2005) demonstrated powerful learning mechanisms that lexical information in NVSS drives improvement of NVSS perception. The study investigated the process of perceptual learning using six-band NVSS sentences while controlling lexical content in the training process (Davis et al., 2005). Listeners were tested using untrained sentences to evaluate the generalized perceptual process for NVSS. Although participants showed less than 10% of word recognition when an NVSS sentence was presented for the first time, their recognition rate increased to 70% after exposure to 30 or 40 NVSS sentences.

Furthermore, perceptual learning of NVSS was enhanced if listeners knew the identity of a target sentence by listening to clear (nondegraded) speech sound or by seeing a written text before listening to the NVSS of the target. In addition, when listeners listened to sentences containing meaningless words during training periods, their performance improvements declined. On the other hand, syntactic prose sentences (i.e., a sentence in

which words were randomly ordered) did not significantly reduce improvement of performance, indicating lexical information in NVSS words facilitates perceptual learning relative to the semantic content of sentences. Perceptual learning was further investigated using monosyllabic and bisyllabic words of six-band NVSS (Hervais-Adelman, Davis, Johnsrude, & Carlyon, 2008). The study assessed that the improvement in listeners' word recognition increased over trials with clear speech presentations before rather than after hearing NVSS. Additionally, the improvements in listeners' performance did not differ between meaningful words and meaningless words, inconsistent with the previous study using sentence materials, indicating the involvement of the sublexical process in perceptual learning (Davis et al. 2005). Thus, it was suggested that the involvement of short-term phonological memory and top-down linguistic processes in the perceptual learning of noise-vocoded speech.

The effect of perceptual learning is known to be sustained for a long time (Karni & Sagi, 1991). For example, even in previous experiments using synthesized speech, the improvements acquired in speech perception were maintained for six months (Schwab, Nusbaum, & Pisoni, 1985) and one year (Altmann & Young, 1993). The long-term maintenance of the improvement of the NVSS perception, however, remains unknown. Furthermore, cochlear implants users need long-term rehabilitation to improve their perception of the new devices. Therefore, the long-term process of training-induced improvements in the NVSS perception may provide further insights for the rehabilitations underlies their daily auditory experiences.

1.2. Neural bases of degraded speech perception

Neural correlates of speech intelligibility

Insights from neuroimaging studies provide the central nervous system for processing NVSS. Early studies measured neural activity associated with intelligibility which indicates comprehensibility of speech signals, by utilizing acoustic characters of speech sounds, including NVSS (Davis & Johnsrude, 2003; Scott et al., 2000). The noise vocoding technique helps control speech intelligibility by varying the number of frequency bands and the cutoff frequency of amplitude envelopes and is employed not only in behavioral experiments but also in neuroimaging studies. These studies using positron emission tomography and functional magnetic resonance imaging (fMRI) describe brain regions whose activation increased with intelligibility and varied across different acoustic forms (spectrally inverted speech; speech segmented by noise bursts and speech in background noise). These studies investigated the cortical regions sensitive to intelligibility and acoustic variation, mainly in the superior temporal gyrus (STG) and superior temporal sulcus (STS). In line with these studies, modulations of intelligibility by utilizing noise vocoding and spectral inversion and manipulations of pitch contour by flatting fundamental frequency were conducted (Kyong et al., 2014). The greater intelligibility was associated with increased activity in the STS in the left hemisphere. The preferential response to the existence of pitch contour was found in the right STG, indicating a right-lateralized process to pitch variation in spoken sentences. Extending their findings, multivariate analyses demonstrated that bilateral STG/STS activation patterns could be classified between intelligible and unintelligible speech (S. Evans et al., 2014; Okada et al., 2010). In addition, speech intelligibility could be discriminated by the

early auditory cortex, Heschl's gyrus (Okada et al., 2010), and by the inferior frontal and inferior parietal cortices (S. Evans et al., 2014).

Neural activity driven by fluctuations in degraded speech comprehension

When listening to severely degraded speech, listeners' comprehension, indicating how much they understand a sentence, fluctuates trial-by-trial, even under a single acoustic clarity condition (Erb, Henry, Eisner, & Obleser, 2013; Erb & Obleser, 2013). In these experiments, listeners were instructed to hear four-band NVSS sentences and repeat the sentences. Erb, Henry, and colleagues (2013) aimed to identify the neural mechanisms of perceptual adaptation. Degraded speech comprehension was then manipulated to improve across trials, while the degree of comprehension fluctuated. The results showed that activity in the thalamus, caudate, frontal, occipital, and cerebellar regions was decreased, and activity in the premotor and posterior cingulate regions was increased gradually as listeners adapted to degrade speech, suggesting cortical and subcortical contributions to adaptation to degraded speech sounds.

Moreover, the results also demonstrated that activity in the bilateral temporal, premotor cortices, and left angular gyrus was correlated positively with the fluctuations in comprehension. Erb and Obleser (2013) assessed the neural activity correlated with fluctuations and compared young and older listener groups. The result showed activity greatly correlated with the fluctuations in the middle frontal gyrus in older listeners relative to young listeners, suggesting a compensatory mechanism in fluctuations in comprehension. These studies revealed that neural activity in the frontal, temporal, parietal, and striatal regions correlated with fluctuation in comprehension rather than

physical stimulus features. However, the monotonic relationship between the degree of comprehension within individuals; however, those findings have not reported a specific activation pattern for a lower or medium degree of comprehension.

Neural correlates of perceptual learning of degraded speech

Perceptual learning of degraded speech has been known to induce neural changes in the auditory and higher cortical regions (Adank & Devlin, 2010; Eisner, McGettigan, Faulkner, Rosen, & Scott, 2010; Hervais-Adelman, Carlyon, Johnsrude, & Davis, 2012; Smalt, Gonzalez-Castillo, Talavage, Pisoni, & Svirsky, 2013; Sohoglu & Davis, 2016). With fMRI measurements, Eisner and colleagues (2010) investigated the neural plasticity associated with perceptual learning of spectrally degraded speech. Participants were trained with degraded speech sentences which were eight-band noise-vocoded and spectrally shifted upward as a simulation of shallow electrode insertions of cochlear implant (Dorman, Loizou, & Rainey, 1997; R V Shannon, Zeng, & Wygonski, 1998). The noise-vocoded and spectrally shifted speech is difficult for naive listeners to comprehend but can be learned with perceptual training, indicating learnable speech. They were also trained with unlearnable speech (spectrally inverted speech) as a control. The fMRI results showed greater responses in the left IFG and STS for learnable stimuli than those for unlearnable stimuli. The left IFG activity varied with individual differences in improvement in degraded speech recognition and with phonological working memory test scores. The left IFG showed increased functional connectivity for learnable speech compared to unlearnable speech with the angular gyrus where showed increased activation as behavioral performance improved across the experimental blocks. These

suggest that interindividual variation in learning ability of the degraded speech is partly associated with language processes in the prefrontal cortex rather than acoustic-phonetic processing in the temporal cortex. In addition to degraded sentences, Hervais-Adelman and colleagues (2012) investigated the learning process of six-band NVSS words. In training sessions, paired presentation of clear then NVSS improved behavioral performance of six-band NVSS recognition throughout experimental blocks. Increased responses during paired clear-then-distorted presentations in the precentral gyrus, which is associated with speech production, suggested a role for the articulatory process of speech in perceptual learning of degraded speech. Similarly, the study investigating the adaptive process of time-compressed speech found neural changes in the auditory temporal regions and premotor cortex accompanying improvement of perceptual performance, suggesting that the learning process depends on remapping novel acoustic patterns onto existing motor network associated with articulatory plans for speech perception (Adank & Devlin, 2010).

Sohoglu and Davis (2016) provided a neurocomputational explanation of perceptual learning of degraded speech with predictive coding theory (Friston, 2005; Rao & Ballard, 1999). The study investigated the commonality between the effect of prior knowledge and the effect of perceptual learning on neural responses to NVSS in the auditory temporal regions with magnetoencephalography (MEG) and electroencephalography (EEG) in order to test whether these phenomena share single mechanisms that minimize prediction errors between sensory inputs and expectations of speech content. The results showed that prior knowledge of NVSS given by written text enhanced subjective clarity of degraded speech and decreased activation in the auditory temporal cortex. Perceptual training improving behavioral performance for NVSS also

reduced activation in the temporal cortex. In addition, improvement in behavioral performance by training was correlated with neural changes induced by perceptual learning and prior knowledge. Together with computational simulations, the study suggested a mechanism based on predictive, interactive processes between sensory processing and higher levels of processing that underpins perceptual learning of degraded speech.

Smalt and colleague (2013) further conducted a more practical, long-term training for two weeks using a real-time CI simulation device. Participants were asked to wear the device implementing an eight-band noise vocoder with an iPod touch and actively listen to speech or music for two hours each day during a two-week training period. Before and after the exposure, changes in behavioral performance and neural activity were measured. The behavioral results showed improvements in degraded word and sentence recognition scores. The results of fMRI also indicated increased responses in the auditory, inferior frontal, supplemental motor, and visual areas, suggesting neural underpinnings for long-term perceptual learning of NVSS. These findings provide insights to investigate neural correlates of perceptual training for cochlear implantation for a long period; however, neural processes of step-by-step improvements and maintenance in training outcome remain poorly understood.

1.3. Hypotheses and objectives

The human central nervous system creates the perception of speech sound from degraded auditory input based on their various support mechanisms. However, it remains unclear how the perceptual performance of degraded speech varies within individuals. From

many neuroimaging evidence, degraded speech perception would be supported by physical acoustic features and several cortical substrates. Therefore, it is hypothesized that involvement of the neural processing in higher auditory and cognitive cortices, such as the STG/STS and ventral frontal cortices, underpins variation in comprehension within individuals. Therefore, the main objective of this thesis work was to characterize variability in the process to degraded speech perception for each individual rather than the process just to stimulus properties like acoustic clarity and external perceptual cues. In so doing, I hope that this research contributes to elucidating neural mechanisms of degraded speech comprehension further.

Study 1: Neural correlates of subjective comprehension of noise-vocoded speech

Given the lack of evidence for cortical mechanisms underlying fluctuations in speech comprehension within individuals, this study examines cortical activation patterns elicited by various degrees of within-individual comprehension of NVSS using fMRI. It reveals that within-individual fluctuations in NVSS comprehension differentially activated the frontotemporal speech-related regions by analyzing activation maps across the whole brain, fMRI signal intensity in regions-of-interest, and asymmetry of activation based on multiscale anatomical labels (lateralization index). In addition, although some neuroimaging research has reported the cortical responses increased monotonically with the degree of comprehension within individuals, those findings have not reported a specific activation pattern for a low or medium degree of comprehension. Thus, the findings allow us to understand variability in the involvement of the frontotemporal regions due to within-individual comprehension.

Study 2: Long-term changes in cortical representation through perceptual learning of spectrally degraded speech

This study aimed to examine the effects of long-term perceptual learning in the auditory modality on human cortical activation. Functional magnetic resonance images were collected while participants performed a perceptual learning task of spectrally degraded speech sound for seven days and the postsession (more than ten months later). The findings indicated that, across experimental days, improvement and persistence in participants' perceptual performance and changes in the neural representation for degraded speech in the left pSTS, associated with the sensory processing of speech signals and connected with multiple sites within the frontal cortex. In addition, commonality in neural representations between the left pSTS and the frontal region emerged through the experimental days. This demonstrates that plastic mechanisms in the frontotemporal cortices underlie the process of perceptual learning of degraded speech.

2. Study 1: Neural correlates of subjective comprehension of noise-vocoded speech

2.1. Introduction

To elucidate the neural mechanisms underlying speech comprehension, studies have examined the neural correlates of speech intelligibility (comprehensibility of a signal based on acoustic clarity; Davis and Johnsrude, 2003; Scott et al., 2000) using distorted speech such as noise-vocoded speech sounds (NVSS), which are intelligible stimuli with reduced spectral detail (Robert V. Shannon et al., 1995). Recent research has reported that comprehension (i.e., how much a listener understand a sentence) of an NVSS sentence within individuals fluctuates trial-by-trial, with a single acoustic clarity condition, and that neural activity in the speech network, including temporal and frontal regions, correlates positively with the level of comprehension (Erb et al., 2013; Erb & Obleser, 2013). These studies showed that neural activity increases monotonically with the degree of comprehension, but the specific neural patterns for each degree of comprehension remain unclear. Neural measures of the variability of comprehension processes within individuals might reveal mechanisms for strategies of listening to degraded speech, and have clinical implications for assessing the efficacy of prosthetic devices and improving hearing rehabilitation. Therefore, it is necessary to understand how neural resources are engaged at different degrees of comprehension.

The superior temporal lobe has been reported to show higher left-lateralization for speech signals with higher acoustic clarity (Kyong et al., 2014; Scott et al., 2000), however, for comprehension within individuals this lateralization could be influenced by

intrinsic higher-level cognitive processing. Activation in the bilateral temporal lobes has been shown to be influenced by lexical processing (Obleser & Kotz, 2010), prior knowledge (Wild, Davis, & Johnsrude, 2012), and attention and listening effort (Davis & Johnsrude, 2003; Wild, Yusuf, et al., 2012). If a listener comprehends words in an NVSS sentence, it could lead to greater recruitment of higher-level processing based on the available linguistic context. Moreover, activation in the left temporal lobe was shown to be suppressed by matching between linguistic cues and sensory inputs, i.e., predictive coding (Blank & Davis, 2016; Sohoglu & Davis, 2016; Sohoglu, Peelle, Carlyon, & Davis, 2012), and therefore high comprehension might also be accompanied by suppression due to intrinsic predictive processes associated with estimation process based on linguistic context. In addition, activity in the left inferior frontal gyrus (IFG) is involved in effortful listening and maximized when speech sounds are moderately distorted, but not when speech sounds are fully intelligible or impossible to understand (Davis & Johnsrude, 2003; Eckert, Teubner-Rhodes, & Vaden, 2016). Together, these higher-level processes could be reflected in lateralized responses in the frontotemporal regions, and be associated with a subjective level of comprehension.

The present fMRI study aimed to investigate which brain regions are involved in the processing of NVSS sentences at different degrees of subjective comprehension reports, and the degree to which their activity is lateralized. We hypothesized that bilateral involvement of temporal cortices would be observed when listeners recognized words in a sentence. We also hypothesized that the left IFG would be recruited for effortful comprehension when they recognized words but did not fully comprehend the sentence.

2.2. Material and methods

Participants

Fifteen right-handed participants participated in the study (mean age: 20.6 ± 0.9 SD; 11 males, 4 females). They were native Japanese speakers with no known hearing impairments, or language or neurological disorders. They were naïve to NVSS. Data from four participants were excluded because they did not report all three degrees of comprehension. Eleven participants were therefore included in the reported analyses (mean age: 21.5 ± 0.8; 8 males, 3 females). All participants gave written informed consent. The experiment was carried out following the experimental protocol approved by Doshisha University Ethics Committee.

Auditory stimuli

The stimulus set consisted of original (non-vocoded) speech sounds, which were always intelligible ("original speech" condition, Fig. 2.1A), noise-vocoded speech sounds ("NVSS" condition, Fig. 2.1B), and spectrally weighted noises, which were always unintelligible and used as a non-speech baseline condition ("noise" condition, Fig. 2.1C). Sixty meaningful Japanese sentences containing 13–16 morae (Japanese phonological syllable-like unit) were spoken by a female speaker and recorded in a sound-proof room, digitized at 44.1 kHz, and re-sampled at 8 kHz. The average duration was 2.4 ± 0.25 s (mean ± SD). Twenty samples of the sentences were used as the original speech set. The other 40 samples were converted into NVSS and noise (20 samples for NVSS and noise; 20 samples for NVSS in behavioral test). The NVSS were generated by applying three-band noise vocoding (0–600, 600–1500, and 1500–4000 Hz; Riquimaroux, 2006). To

observe different degrees of comprehension within individuals, we avoided using four-band NVSS, which are nearly perfectly intelligible (Ueda et al., 2018) and lead to a dramatic improvement of speech comprehension (Erb et al., 2013; Riquimaroux, 2006), and instead used three-band NVSS as more degraded stimuli with relatively low intelligibility and learnability. A band noise in each frequency range was generated by multiplying band-passed white noise with amplitude envelopes obtained from low-pass-filtered recorded speech sound at 16 Hz. The spectrally weighted noises were composed of three band-passed white noises with an average amplitude matching the corresponding band in recorded speech. Unlike the NVSS, the amplitude envelope was not modulated using recorded speech. The average duration of the noises was 2.5 ± 0.25 s (mean ± SD). To avoid perceptual learning of NVSS, the same stimulus was never repeatedly presented to any participant, and feedback (written text or original speech version of NVSS) was never given to participants.

Behavioral test

The comprehension performance of NVSS for each participant was measured before and after the MRI scan. To avoid adaptation to each NVSS sentence, stimuli were not reused across the behavioral test and the fMRI sessions. In each test, ten NVSS stimuli were presented through headphones (ATH-A900, Audio-Technica Corp., Tokyo, Japan) in a sound-proof room. The stimuli were played at a comfortable listening volume adjusted before the beginning of the test. Participants were instructed to listen carefully to NVSS and write down what they heard.

Functional MRI procedure

An event-related fMRI run was conducted in which 20 NVSS trials, 20 original speech trials, and 20 noise trials were presented in a pseudorandom order. The stimuli were presented through MRI-compatible headphones (AS3000K, Kiyohara Optics Inc., Tokyo, Japan), at a comfortable level adjusted before the experiment began. After every 10 trials, a 9-s rest was presented. The total duration of the fMRI run was 11 min. All On each trial, the sound stimulus was accompanied by a white fixation cross shown on a black screen (Fig. 2.1D). Participants were asked to listen carefully and comprehend the sound stimuli while focusing on the cross. To evaluate trial-by-trial comprehension within individuals, participants were asked to press one of the three buttons on a response box to report the degree of comprehension (“high”: whole sentence, “medium”: at least one word, “low”: no words).

MRI acquisition and data analyses

Functional images were acquired (gradient-echo echo-planar imaging; TR = 9.0 s; TE = 50 ms; 25 slices; 4-mm thickness with a 1-mm gap; transverse; FA = 90°; FOV = 192 × 192 mm; matrix size = 64 × 64) on a 1.5-Tesla MRI scanner (Echelon Vega, Hitachi Medical, Chiba, Japan) using the sparse sampling method (Hall et al., 1999), where sound stimuli were presented during silent intervals between scans to reduce the effect of scanner noise (Fig. 2.1D). The first volume of the run was discarded to allow magnetization to reach equilibrium, resulting in 72 volumes available for analysis. T1-weighted structural images were also acquired for anatomical reference.

Data analysis was performed using SPM12 software (Wellcome Department of

Imaging Neuroscience, London, UK; http://www.fil.ion.ucl.ac.uk/spm/). Images were realigned, spatially normalized to an echo-planar imaging template in Montreal Neurological Institute (MNI) space, and smoothed with a 10-mm full-width at half-maximum Gaussian kernel. The single-subject level statistical analyses were then conducted using a general linear model. Blood-oxygen-level-dependent (BOLD) signals for three conditions (NVSS, original speech, and noise) and five conditions (NVSS with low/medium/high comprehension, original speech, and noise) were modeled with a hemodynamic response function. At the group level, random effect analyses were conducted by entering the single-subject level activation data. Activation maps were thresholded at $p < 0.001$ at the voxel level and corrected for the family-wise error at the cluster level, threshold $p < 0.05$, for multiple comparisons across the whole brain. We identified the cluster-level threshold using a nonparametric permutation test with the SnPM toolbox (http://warwick.ac.uk/snpm) to reduce the false positive rate associated with parametric statistics (Eklund, Nichols, & Knutsson, 2016). Estimations were based on 5000 permutations. Additionally, Harvard–Oxford atlas-based region-of-interest (ROI) analysis was performed within the regions associated with sentence context (Obleser & Kotz, 2010) using MarsBaR (http://marsbar.sourceforge.net/). Differences in BOLD signals between the conditions were compared using nonparametric Wilcoxon signed-rank tests to account for the small sample size (n = 11) and avoid assuming normality.

To assess whether comprehension differently influences on the lateralization, a lateralization index (LI) was calculated using a well-established bootstrap procedure implemented within the LI-toolbox (Wilke & Lidzba, 2007). LI is calculated using the formula:

$$LI = (\Sigma activation_{left} - \Sigma activation_{right})/(\Sigma activation_{left} + \Sigma activation_{right}). \quad [1]$$

Σactivation refers to the sum of the voxel values of the contrasts (NVSS with each degree of comprehension > noise and original speech > noise) of a group-level activation map within an anatomical mask. To avoid the problem of statistical thresholding dependency (Bradshaw, Bishop, & Woodhead, 2017), LI was calculated at multiple statistical thresholds. We used weighted mean LI, which is calculated by giving greater weighting to LIs at higher thresholds. LI > 0.2 and LI < −0.2 have been widely interpreted as left-lateralization and right lateralization, respectively (e.g., Kyong et al., 2014). Considering regional heterogeneity in lateralization (Bradshaw et al., 2017), we analyzed lateralization at multiple scales (the cerebral lobes and the Harvard–Oxford atlas regions).

In addition, to test the effect of adaptation (i.e., performance improvement) on neural activity (Erb et al., 2013), we compared changes in BOLD signal across trials between NVSS and original speech, which were modeled by two parametric modulators. We modeled a parametric modulator for NVSS by multiplying trial numbers with estimated perceptual performances, resulting in a quadratic curve. Each participant's estimated performance for NVSS was linearly increased from their pre-scan to their post-scan test score. A parametric modulator for original speech was represented as trial numbers, resulting in a linear curve, because there were no changes in performance for original speech.

2.3. Results and discussion

Behavioral performance

The behavioral comprehension performance was scored as the proportion of correctly

understood mora of NVSS sentences. The mean performance across participants was 36.3 ± 7.6% (mean ± SD) on average across pre-scan and post-scan tests (pre-scan test: 31.7 ± 10.5%, post-scan test: 40.8 ± 7.2%). The difference between pre-scan and post-scan tests (9.1 points; $t(10) = 3.09$, $p < 0.05$) suggests there was a learning effect; however, this was not the focus of this study. In addition, there was no effect of stimulus type (NVSS, clear speech, and noise; $F(1,10) = 2.19$, $p = 0.17$) and no effect of comprehension degree ($F(1,10) = 2.09$; $p = 0.18$) on reaction times (Fig. 2.2) in the fMRI experiment.

Cortical activity during speech comprehension

In the fMRI experiment, the degree of subjective comprehension was collected with the trial-by-trial comprehension report. The rates of each degree of comprehension were 40.0 ± 18.7% for low, 48.2 ± 17.1% for medium, and 11.8 ± 5.6% for high comprehension. We assessed the neural activity involved in listening to either NVSS or original speech contrasted with that during noise stimuli (Fig. 2.3A, Table 2.1). The original speech activated the bilateral superior temporal gyrus (STG), whereas NVSS showed additional activity in the left middle temporal gyrus (MTG) and left IFG, consistent with fMRI studies reporting activity for noise-vocoded sentences compared with unintelligible control stimuli (Davis & Johnsrude, 2003). We then examined the activity evoked by NVSS with each degree of comprehension as compared with noise (Fig. 2.3B, Table 1). NVSS with low comprehension activated the posterior to anterior area of the left STG (Fig. 3B, blue). The left STG activity was commonly observed for all three degrees of comprehension of NVSS (Fig. 2.3B, white). Additionally, medium comprehension of NVSS showed activity in the left medial frontal gyrus, right STG and left IFG (Fig. 2.3B,

green). High comprehension of NVSS induced activity in the left posterior MTG and temporal pole (Fig. 2.3B, red). The right STG showed stronger activation for NVSS with higher comprehension, which is concordant with the observation that comprehension is linked with linearly increased activity (Erb et al., 2013).

In addition, the ROI analysis (Fig. 2.3C) revealed significantly greater activation, compared with noise, in the left inferior frontal gyrus, pars opercularis (IFGop) for NVSS with medium comprehension, in the left posterior STG for original speech and all three degrees of NVSS, and in the right posterior STG for original speech and medium and high comprehension of NVSS. Moreover, the right posterior STG also showed greater signals for high comprehension than for medium comprehension of NVSS, indicating a preferential response to high comprehension within a single intelligibility condition.

LI analysis demonstrated that in the temporal lobe (Figs. 2.4A and B) NVSS with low comprehension showed left lateralization, but bilateral activation was observed for NVSS with medium/high comprehension (LI for low: 0.76; medium: 0.02; high: 0.11). Further, using atlases of the primary and higher auditory cortex (Figs. 2.4A and B), all degrees of NVSS comprehension were left lateralized in Heschl's gyrus (low: 0.68; medium: 0.43; high: 0.55). By contrast, higher comprehension reduced left-lateralization in the posterior STG (low: 0.76; medium: –0.03; high: –0.37).

Furthermore, additional exploratory LI analyses were conducted into other anatomical ROIs covering the whole cerebral cortex. In the whole hemisphere ROI (Figs. 2.5A, B), NVSS with low/medium comprehension was left lateralized but high comprehension was not (LI for low: 0.73; medium: 0.35; high: 0.12), while original speech was left lateralized (0.48). When subdividing into cortical lobes (Figs. 2.5A, C), in addition to the temporal lobe, the frontal lobe showed left-lateralization regardless of

comprehension degree, and demonstrated a relatively greater LI for medium comprehension than low and high comprehension (low: 0.34; medium: 0.70; high: 0.23). Using more fine-grained atlases in the temporal regions (Figs. 2.5D, E), all degrees of NVSS comprehension were left lateralized in the temporo-occipital part of MTG (low: 0.87; medium: 0.88; high: 0.84) in addition to the Heschl's gyrus. By contrast, higher comprehension reduced left-lateralization in the anterior STG (low: 0.21; medium: -0.15; high: -0.21), the posterior STG (low: 0.76; medium: -0.03; high: -0.37), the anterior MTG (low: 0.46; medium: 0.02; high: -0.50), and the posterior MTG (low: 0.61; medium: -0.07; high: -0.70).

To examine the effect of performance improvement on the frontal activation (Erb et al., 2013), we analyzed activation associated with adaptation to NVSS compared with original speech (see Material and methods). The results did not show significant activation in the frontal regions even at a lower threshold ($p < 0.01$, uncorrected for multiple comparisons). This could be due to relatively small performance improvement in this experiment, and suggests that improvement did not significantly affect activation associated with comprehensions.

Our results suggest that variation in comprehensions within individuals can be reflected in asymmetric activity in the frontotemporal speech network (Figs. 2.3B and 2.4B). Although previous studies revealed that cortical activations for degraded speech are modulated by stimulus properties such as acoustic clarity (Davis & Johnsrude, 2003; Scott et al., 2000) and sentence context (Obleser & Kotz, 2010), we found activation associated with variability in speech comprehension within a single intelligibility condition, by using trial-by-trial variations of subjective comprehension reports. In particular, lateralization analysis provided evidence of comprehension-related changes in

lateralization in the temporal cortices.

The temporal lobe showed less lateralized responses to medium and high comprehension of NVSS, while low comprehension of NVSS and original speech, which was fully comprehended, led to a left-lateralization (Figs. 2.3B). Although we did not observe direct evidence, the lateralization of higher comprehension of NVSS might reflect the involvement of listening effort (Eckert et al., 2016; Wild, Yusuf, et al., 2012) and linguistic processes (Obleser & Kotz, 2010) that mediate degraded speech comprehension.

The ROI analysis in the STG (Fig. 2.3C) showed that comprehension of NVSS significantly increased activity in the right hemisphere but not in the left, even though higher comprehension accompanying higher-order processes might have been expected to increase activation bilaterally in the STG. This observation may be due to suppression of activity in the left STG induced by predictive process for speech comprehension (Sohoglu et al., 2012), meaning the BOLD signal intensity in the left hemisphere did not increase with comprehension. In contrast, the right STG showed increased activity in the high relative to the medium comprehension condition. The right temporal lobe has been associated with processing of suprasegmental prosody, such as accents and tones, and melodic information (Kyong et al., 2014; Meyer, Alter, Friederici, Lohmann, & Cramon, 2002; Poeppel, 2003). Although the fundamental frequency information is not available in NVSS, previous studies revealed that suprasegmental prosody can be successfully recognized in NVSS words (Fu et al., 1998; Riquimaroux, 2006; Xu et al., 2002). Hence, it is possible that greater recognition of words and sentences might be associated with greater recognition of linguistic prosody, and that the involvement of the right temporal lobe might be increased with higher comprehension of NVSS. In addition, a meta-analysis

found that the right temporal regions show involvement in sentence context processing (Vigneau et al., 2011) and might be associated with context processing in compensation for speech degradation.

Although we did not find significant differences in activation of the left IFG between degrees of comprehension, the left IFG showed greater activation in medium comprehension of NVSS condition than noise condition (Figs. 2.3B and C), suggesting that the left IFG might be required for subjectively moderate difficulty NVSS. Word recognition without full comprehension of an NVSS sentence (i.e., medium comprehension) might cause ambiguous semantic context. The partial comprehension of a sentence might lead to greater processing demands of semantic context (Obleser & Kotz, 2010) and listening effort (Davis & Johnsrude, 2003; Eckert et al., 2016; Wild, Yusuf, et al., 2012) in the left IFG.

There are a few limitations in this study. The sample size was relatively small, and may negatively affect statistical power, although nonparametric tests were used to avoid normality assumption. In addition, the number of comprehension responses of NVSS was unbalanced, especially for high comprehension. Further studies with a large number of participants and more appropriate difficulty of stimuli for balanced responses are needed to replicate and confirm our results.

In summary, our results suggest that the hemispheric specialization might indicate responses to the variability in speech comprehension under an acoustically degraded condition. While low comprehension of NVSS was reflected in a left-lateralization in the temporal cortex (Fig. 2.4B), greater involvement of the bilateral temporal regions and the left IFG was observed when a participant recognized words in an NVSS (Figs. 2.3B, C and 2.4B). Sentence level comprehension was more associated with right-lateralized

responses in the STG (Fig. 2.4B). This study focused on the degree of subjective comprehension, however, we did not measure actual speech comprehension. Investigations of neural correlates of both subjective and actual comprehension would further reveal how the processes associated with subjective comprehension are linked to successes and errors in the degraded speech understanding and shed more light on the neural underpinnings of variability in speech comprehension.

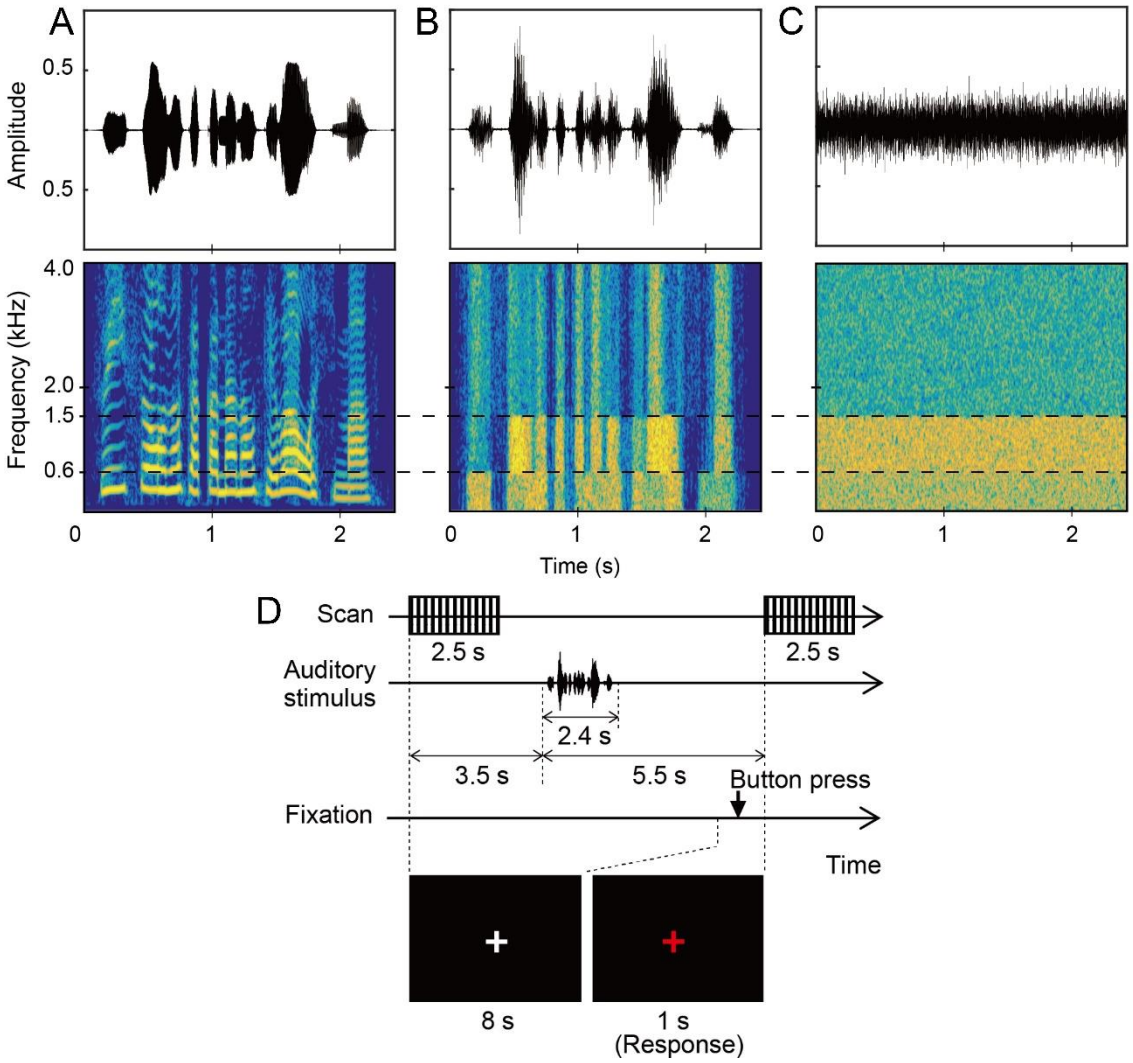

Fig. 2.1. Examples of sound stimuli /suisô no sakana no sewa wo shimasu/ ("I take care of fish in a tank.") in original speech (A), NVSS (B), and weighted noise (C). Upper panels: temporal amplitude patterns. Lower panels: sound spectrograms. (D) Temporal structure of a single trial; scanning, presentation of auditory stimuli, and eye fixations with a button press. An auditory stimulus was presented during a silent gap between scans to avoid noise generated by the scanning. A white fixation cross was shown for 8 s on the black screen, and a red fixation cross was displayed for the last second, instructing the participant to press a button for a comprehension report.

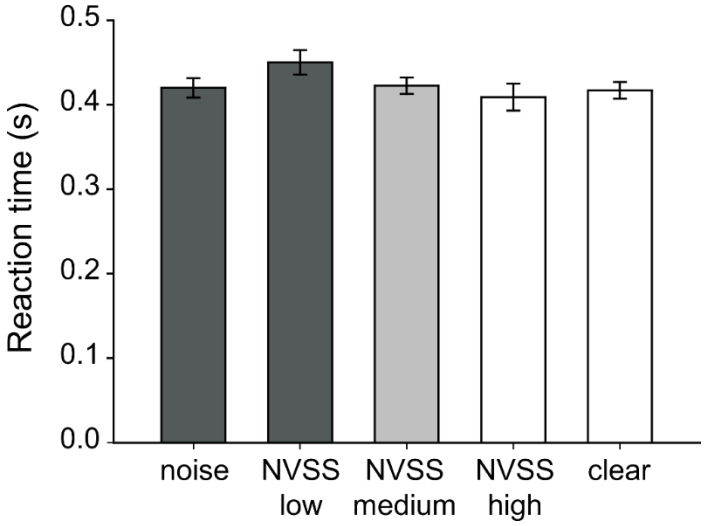

Fig. 2.2. Reaction times in the fMRI experiment. Error bars indicate standard error of the mean. low = low comprehension; medium = medium comprehension; high = high comprehension.

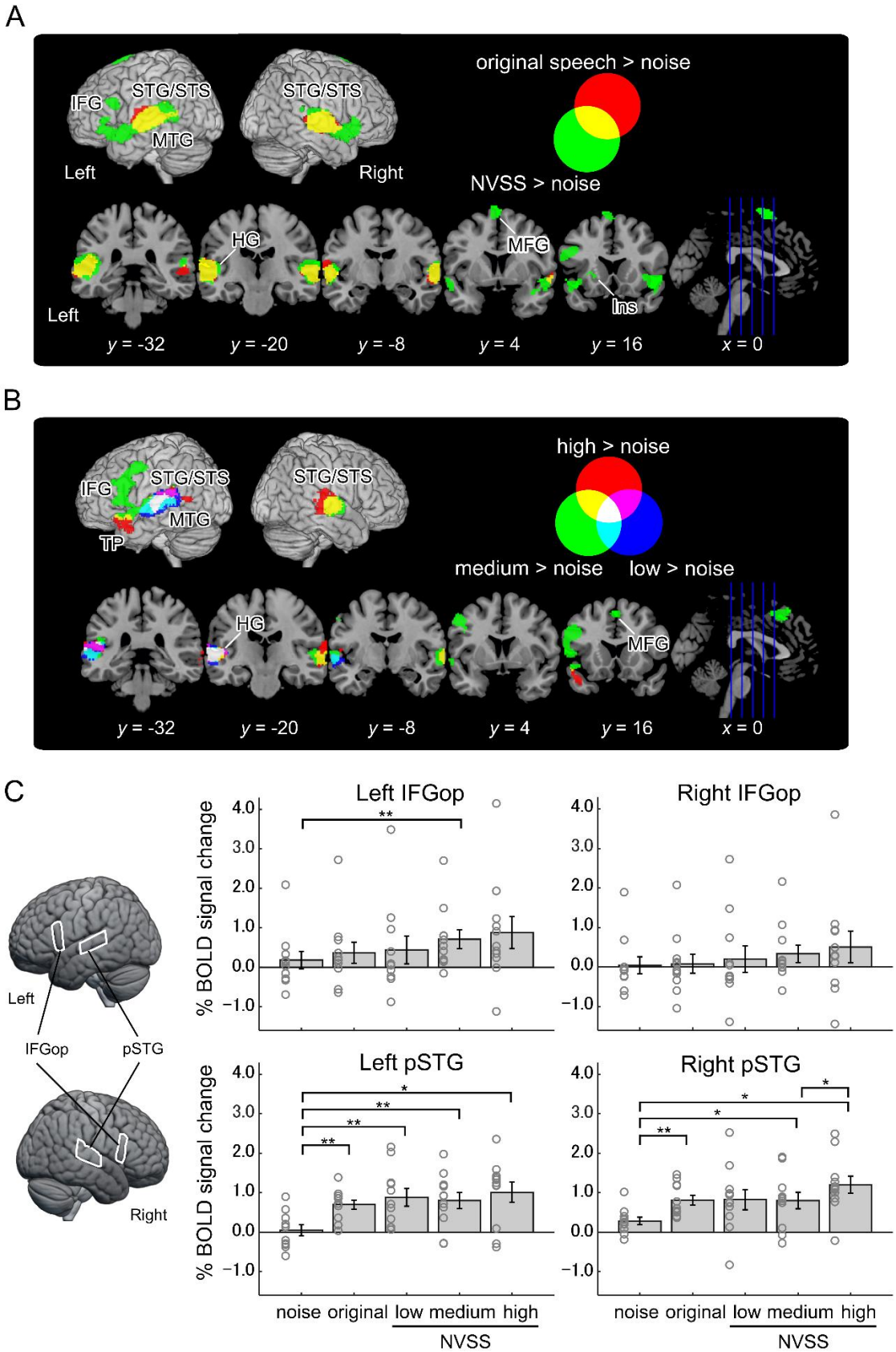
A
original speech > noise
IFG
STG/STS
MTG
STG/STS
Left
Right
NVSS > noise
HG
MFG
Ins
Left
y = -32
y = -20
y = -8
y = 4
y = 16
x = 0
B
high > noise
IFG
STG/STS
MTG
TP
STG/STS
medium > noise
low > noise
HG
MFG
y = -32
y = -20
y = -8
y = 4
y = 16
x = 0
C
Left
IFGop
pSTG
Right
Left IFGop
Right IFGop
Left pSTG
Right pSTG
% BOLD signal change
4.0
3.0
2.0
1.0
0.0
−1.0
noise original low medium high
NVSS

Fig. 2.3. fMRI group results. (A) Activated areas for NVSS minus noise (green) and original speech minus noise (red). The overlap between the two activated areas is in yellow, depicted in the Venn diagram. (B) Activated areas for NVSS with high comprehension minus noise (red), NVSS with medium comprehension minus noise (green), and NVSS with low comprehension minus noise (blue). The overlap between the activated areas is in yellow, light blue, pink, and white, as depicted in the Venn diagram. (C) Results from ROI analysis in the bilateral inferior frontal gyrus, pars opercularis (IFGop), and posterior superior temporal gyrus (pSTG). Left: ROI locations displayed on a template brain. Right: The group mean percent BOLD signal change in each ROI. Gray circles represent individual data. The significant differences between conditions are displayed by the asterisks (*$p < 0.05$, **$p < 0.01$; two-sided Wilcoxon signed-rank tests after Bonferroni–Holm correction). *x*, *y* = MNI-coordinates; IFG = inferior frontal gyrus; STG = superior temporal gyrus; STS = superior temporal sulcus; MTG = middle temporal gyrus; MFG = medial frontal gyrus; Ins = insula; TP = temporal pole; low = low comprehension; medium = medium comprehension; high = high comprehension.

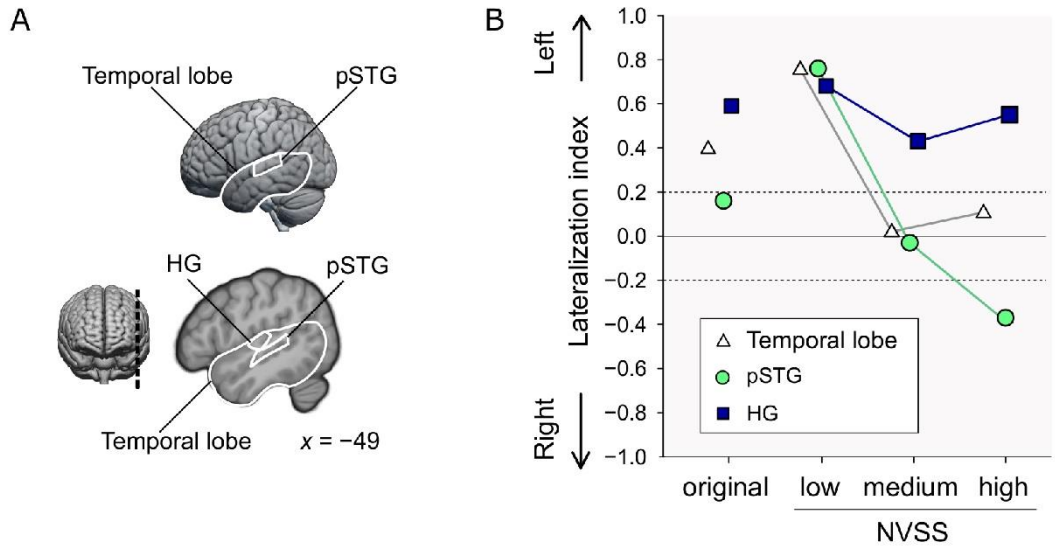

Fig. 2.4. Results of lateralization index (LI) analysis. LIs were calculated using activation maps resulting from group analyses for the following contrasts: original speech minus noise, NVSS with high comprehension minus noise, NVSS with medium comprehension minus noise, and NVSS with low comprehension minus noise. (A) Locations of anatomical masks displayed on the left hemisphere of a template brain. The locations are also displayed on the sagittal section to show the primary auditory cortex, i.e., Heschl's gyrus. (B) Plots of the LIs of the temporal lobe and Harvard-Oxford atlas regions within the temporal cortex. pSTG = posterior superior temporal gyrus; HG = Heschl's gyrus; *x* = MNI coordinate.

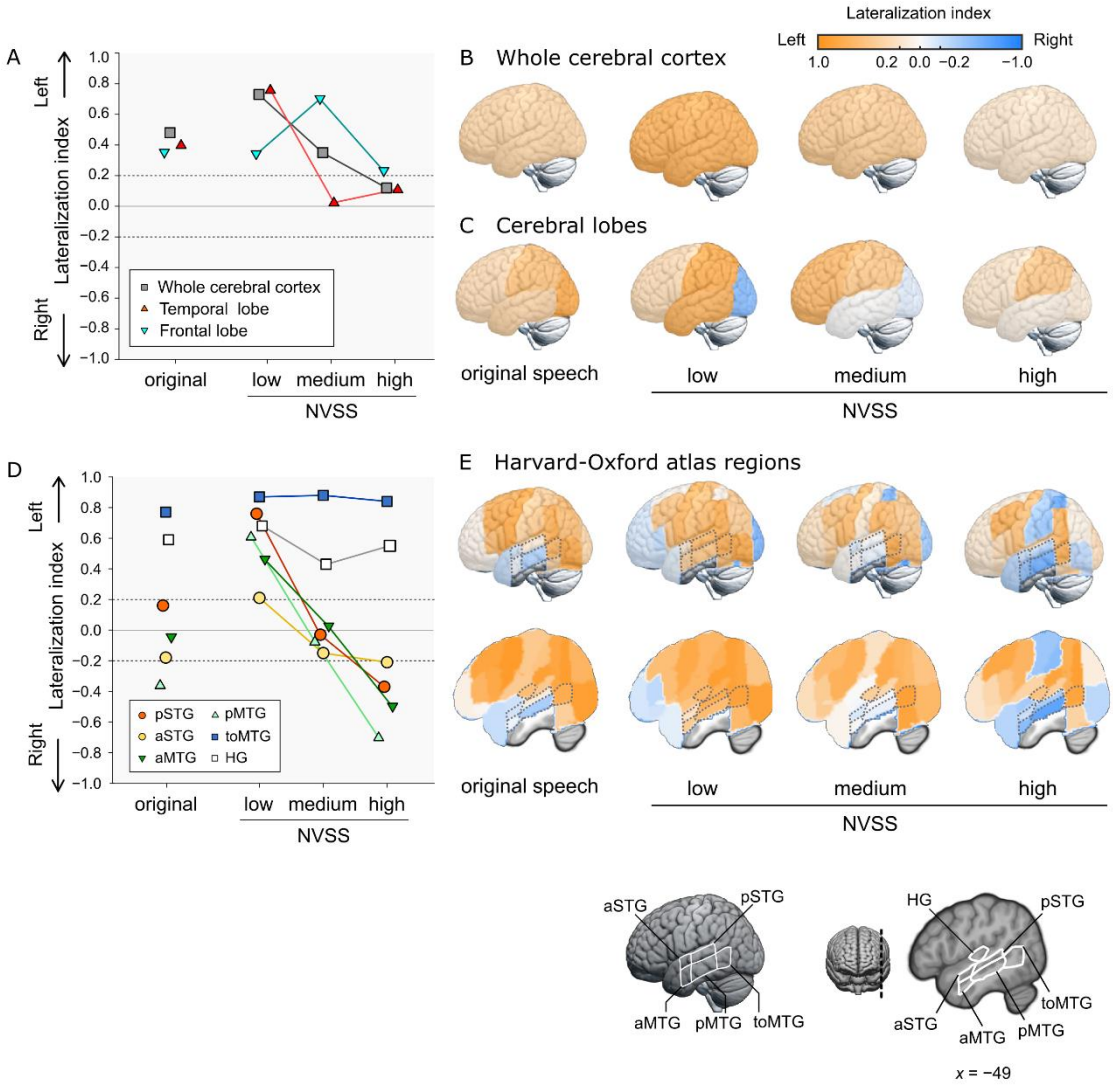

Fig. 2.5. Summary of lateralization index (LI) on multiple anatomical atlas. LIs were calculated for the following contrasts: original speech minus noise, NVSS with high comprehension minus noise, NVSS with medium comprehension minus noise, and NVSS with low comprehension minus noise. (A) Plots of the LIs of the whole cerebral cortex and cerebral lobes, specifically the temporal and frontal lobes. (B and C) Orange-blue represents the LI of the whole cerebral cortex (B), and the cerebral lobes rendered on the left hemisphere of a template brain. (D) Plots of the LIs of Harvard-Oxford atlas regions within the temporal cortex. (E) Orange-blue represents the LI of Harvard-Oxford

atlas regions (E, upper row) rendered on the left hemisphere of a template brain. LIs of Harvard-Oxford atlas regions are also displayed on the sagittal section to show the primary auditory cortex, i.e., Heschl's gyrus (E, lower row). *x* = MNI coordinate.

Table 2.1 Clusters of brain regions that showed significant activation in the whole-brain analysis labeled according to SPM Anatomy Toolbox.

Contrast	Anatomical label of cluster peaks	Peak location in MNI coordinates (mm)			Cluster size (Voxels)	*t* value
		x	*y*	*z*		
original > noise						
	L Superior Temporal Gyrus	−60	−20	10	2180	15.50
	L Superior Temporal Gyrus	−58	−14	0		14.64
	L Superior Temporal Gyrus	−50	−24	10		10.65
	R Superior Temporal Gyrus	66	−4	0	1213	13.30
	R Superior Temporal Gyrus	58	−14	6		9.69
	R Superior Temporal Gyrus	68	−22	0		9.17
NVSS > noise						
	L Superior Temporal Gyrus	−62	−22	10	3996	22.06
	L Superior Temporal Gyrus	−52	−34	14		12.56
	L Middle Temporal Gyrus	−54	−32	6		12.01
	R Superior Temporal Gyrus	68	−14	2	2722	15.59
	R Superior Temporal Gyrus	62	−18	10		11.54
	R Insula Lobe	50	8	−6		9.98
	L Posterior Medial Frontal	−4	10	68	445	13.31
	L Inferior Frontal Gyrus (pars Triangularis)	−50	18	24	365	7.12
NVSS with low comprehension > noise						
	L Middle Temporal Gyrus	−66	−22	2	1783	10.73
	L Middle Temporal Gyrus	−64	−36	4		8.32
	L Superior Temporal Gyrus	−56	−38	14		7.51
NVSS with medium comprehension > noise						
	R Superior Temporal Gyrus	66	−6	2	607	11.11
	R Superior Temporal Gyrus	68	−14	−6		7.94
	R Superior Temporal Gyrus	58	−18	4		5.52

L Inferior Frontal Gyrus (pars Triangularis)	−50	18	28	3142	8.69
L Superior Temporal Gyrus	−60	−20	6		8.13
L Inferior Frontal Gyrus (pars Orbitalis)	−46	24	−10		7.75
R Superior Medial Frontal Gyrus	6	30	56	484	6.23
L Posterior Medial Frontal	0	18	52		5.86
L Superior Medial Frontal Gyrus	−2	32	56		5.63
NVSS with high comprehension > noise					
R Superior Temporal Gyrus	68	−12	0	575	8.26
R Superior Temporal Gyrus	70	−18	8		7.86
R Rolandic Operculum	70	−20	16		6.53
L Superior Temporal Gyrus	−52	−18	8	1003	7.86
L Superior Temporal Gyrus	−56	−36	16		6.60
L Middle Temporal Gyrus	−66	−16	0		6.53
L Temporal Pole	−42	20	−28	301	7.17
L Medial Temporal Pole	−50	10	−28		6.88
L Temporal Pole	−44	24	−20		5.85

3. Study 2: Long-term changes in cortical representation through perceptual learning of spectrally degraded speech

3.1. Introduction

Speech perception system adapts to acoustically degraded conditions via training, and improvement in perceptual performance is sustained for long periods. Recipients of cochlear implants, who receive spectrally degraded inputs via electrodes by stimulating auditory nerves, improve their perception with long-term rehabilitations after implantation, and the improvement underlies their daily auditory experiences. Behavioral experiments using artificially degraded sounds with normal hearing participants have been conducted to provide insights for mechanisms of perceptual learning of degraded speech and exhibited training-induced improvements in the perception (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005). Functional neuroimaging studies have explored to understand how the perceptual learning of degraded speech has been underpinned by auditory and higher-level processing, associated with linguistic, working memory, and articulatory processes (Adank & Devlin, 2010; Davis et al., 2005; Eisner et al., 2010; Hervais-Adelman et al., 2012; Sohoglu & Davis, 2016). Focusing further on neural processes for improvement and maintenance in long-term learning in degraded speech perception would advance our knowledge of adaptive mechanisms of perceptual systems and contribute training design for rehabilitation of auditory disorders.

Using consecutive training experiments within a day, changes in neural activations across multiple trials and sessions were observed in the left posterior superior temporal sulcus (pSTS), inferior frontal gyrus (IFG), ventral premotor cortex (vPM), and parietal regions (Adank & Devlin, 2010; Eisner et al., 2010; Erb et al., 2013) accompanying improvement of perceptual performance of degraded speech. The left pSTS is a regions associated with sensory processing of speech signals and connected with multiple site within the frontal cortex (Hickok & Poeppel, 2007; Scott & Johnsrude, 2003). Smalt et al. (2013) conducted a long-term training for two weeks and measured activation changes comparing two sessions before and after training, demonstrating more activations after training were observed in the temporal regions including the left pSTS. The left IFG has been associated with effortful listening (Davis & Johnsrude, 2003), semantic context (Obleser & Kotz, 2010), and phonological working memory (Eisner et al., 2010) for processing of degraded speech. In addition, the functional connectivity between the left IFG and the left pSTS has been observed during training for degraded speech perception (Eisner et al., 2010). The left vPM showed neural representations of heard syllables of degraded speech, suggesting involvement of perceptual processing based on articulatory information (Samuel Evans & Davis, 2015). Activity in the left vPM was also suggested to enhance perceptual learning of degraded speech (Hervais-Adelman et al., 2012).

With regard to long-term observations of improvements in perceptual performances, visual perceptual learning studies showed that changes in neural activity across multiple days have been observed in the visual perception-related sensory cortices (Frank, Reavis, Tse, & Greenlee, 2014; Yotsumoto, Watanabe, & Sasaki, 2008), and maintenances of the neural changes was observed a few weeks or years after training ended have been also demonstrated (Aloufi, Rowe, & Meyer, 2021; Bi, Chen, Zhou, He, & Fang, 2014; Chen,

Lu, Shao, Weng, & Fang, 2017; Frank, Greenlee, & Tse, 2018; Frank et al., 2014). In addition to changes in the sensory cortices themselves, enhanced connectivity between the sensory and higher-order cortices has been shown (Chen et al., 2015, 2017; Frank et al., 2014). Although these cortical changes have been investigated with the different modality, these studies suggested that perceptual learning of speech perception would also be underpinned by the long-term processes of plastic changes in cortical regions associated with sensory and higher-order processing. However, neural processes of step-by-step improvements across days and of maintenance in speech perception remain poorly understood.

The aim of this study is to demonstrate time courses of consecutive changes and maintenance of neural activation pattern accompanying improvements in perceptual performances to extend our understanding of neural basis of long-term learning mechanisms of speech perception. We conducted the functional magnetic resonance imaging (fMRI) experiments in which normal hearing participants were trained to comprehend noise-vocoded speech sounds (Robert V. Shannon et al., 1995) as a simulation of the cochlear implant, referred to as NVSS. We then measured neural activity during listening to NVSS with fMRI for seven experimental days. In addition, given that the training-induced improvements once acquired in degraded speech perception were maintained for one year (Altmann & Young, 1993), we conducted the experiments with the same participants approximately one year later to observe long-term effects of behavioral and neural changes. We used the representational similarity analysis (RSA; Kriegeskorte et al. 2008) to assess neural representations of NVSS in the pSTS, IFG, and vPM. The RSA is a multivariate approach in which dissimilarity distances were computed between the neural activation pattern associated with NVSS and acoustically clear speech

(fully intelligible), to rule out a potential confound of familiarity for the task due to a long-term period of the experiment and a potential confound of linguistic processing induced by successful speech perception. We then evaluated learning-dependent changes in the activation patterns for NVSS and clear speech across experimental days due to clarify differences in the process of speech perception. We hypothesized that the sensory and higher-order regions such as the pSTS, IFG, and vPM displayed changes in activation patterns for NVSS during experimental days, while participants' perceptual performance increased. Moreover, we hypothesized that the once acquired perception and neural change for NVSS in pSTS were maintained for a long period.

3.2. Material and methods

Participants

Five right-handed participants took part in the study (mean age: 21.4 ± 1.5 SD; four males, one females). They were native Japanese speakers with no known hearing impairments, or language or neurological disorders. All participants gave a written informed consent. They were naïve to NVSS. The experiment was carried out following the experimental protocol approved by Doshisha University Ethics Committee.

Auditory stimuli

The stimulus set consisted of non-vocoded speech sounds, which were always intelligible ("clear speech" condition) and noise-vocoded speech sounds ("NVSS" condition). Two hundred and sixty meaningful Japanese sentences containing 13–16 morae (Japanese phonological syllable-like units) were spoken by a female speaker and recorded in a

sound-proof room, digitized at 44.1 kHz, and re-sampled at 8 kHz. The average duration was 2.4 ± 0.2 s (mean ± SD). Eighty samples of the sentences were used as the clear speech set. The other 180 samples were converted into NVSS (80 samples for fMRI experiments; 100 samples for behavioral tests). The NVSS were generated by applying three-band noise vocoding (0–600, 600–1500, and 1500–4000 Hz; Riquimaroux 2006; Shannon et al. 1995). A band noise in each frequency range was generated by multiplying band-passed white noise with an amplitude envelope within each frequency range extracted from recorded speech. The amplitude envelope was generated by half-wave rectification followed by low-pass-filtering sound at 16 Hz. Sentences were randomly allocated to NVSS and clear speech. They were presented in a randomized order per participant. The same sentence was never used for clear speech and NVSS across trials within a participant.

Experimental procedures

One behavioral test experiment and two fMRI experiments were conducted on each day (Fig. 3.1A). The experiments were performed for seven days within a nine-day period. More than 10 month (mean = 390 days, SD = 78 days, range = 325–488 days) after the end of the seven experiment days, both behavioral and fMRI experiments were further performed as a post session. In the behavioral test, the comprehension performance of NVSS for each participant was measured immediately after the MRI experiments. Only in the Day 1 and the post sessions, the performance was also measured before the MRI experiments. In each test, ten NVSS stimuli not used in the fMRI study were presented through headphones (ATH-A900, Audio-Technica Corp., Tokyo, Japan) in a sound-proof room. Participants were instructed to listen carefully to NVSS and write down what they

heard. The behavioral performance was scored as the proportion of correctly recognized morae of NVSS sentences.

The fMRI experiment consisted of an NVSS session and a clear speech session on each day. The NVSS session consisted of ten trials. Adapting the sparse sampling method (Hall et al., 1999), sound stimuli were presented during silent intervals between fMRI scans to reduce the effect of scanner noise (Fig. 3.1B). On each trial, nine seconds after a trial start a notification sound was presented, four repetitions of a randomly selected NVSS sentence were presented every nine seconds. The first two NVSS presentations were accompanied with a fixation cross shown on a screen, and then the second two NVSS presentations were accompanied with a written presentation of the sentence on the screen. In the first two NVSS presentations, participants were instructed to listen to NVSS carefully and comprehend the sound stimuli while focusing on the cross. In the next two presentations, participants were required to comprehend NVSS while viewing the sentence on the screen. After that, participants were instructed to push a response button whether they were able to correctly perceived NVSS in the first two NVSS presentations. In the clear speech session, instead of NVSS, a randomly selected clear speech sentence was presented. These sound stimuli were presented through MRI-compatible headphones (AS3000K, Kiyohara Optics Inc., Tokyo, Japan) in the scanner.

MRI acquisition and data analyses

Functional images were acquired (gradient-echo echo-planar imaging; TR = 9.0 s; TA = 2.5 s; TE = 50 ms; 25 slices; 4-mm thickness with a 1-mm gap; transverse; FA = 90°; FOV = 192 × 192 mm; matrix size = 64 × 64) on a 1.5-Tesla MRI scanner (Echelon Vega, Hitachi Medical, Chiba, Japan) using the sparse sampling method (Hall et al., 1999).

T1-weighted structural images were also acquired for anatomical references.

Data analysis was performed using SPM12 software (Wellcome Department of Imaging Neuroscience, London, UK; http://www.fil.ion.ucl.ac.uk/spm/). Images were realigned, spatially normalized to an echo-planar imaging template in Montreal Neurological Institute (MNI) space, and smoothed with a 10-mm full-width at half-maximum Gaussian kernel. The single-subject level statistical analyses were then conducted using a general linear model. Neural activity (BOLD signals) for each trial in NVSS and clear speech condition was modeled with a hemodynamic response function. Head movement parameters from realignment correction were added as regressors of no interest. The resulting beta values (i.e., estimates of neural activity) were obtained per trial, voxel and participant.

The RSA was conducted using CoSMoMVPA toolbox (Oosterhof, Connolly, & Haxby, 2016) to analyze neural dissimilarity between functional responses evoked by NVSS and clear speech (Figs. 3.1C and D). We made a vector representing beta values of voxels for each trial, region-of-interest (ROI), and participant (Fig. 3.1B). The beta values were demeaned by subtracting the mean beta value across all trials within the same day per each individual. We then computed dissimilarity index expressed by correlation distance ($1 - \rho$; ρ indicates spearman correlation coefficient) between the 20 vectors (10 vectors for the NVSS conditions and 10 vectors for the clear speech conditions) to create a neural representational dissimilarity matrix (RDM), resulting in a 20×20 RDM for each day (Fig. 3.1C). Next, we defined a simple model RDM representing stimulus types i.e., NVSS and clear speech (Fig. 3.1D right). Finally, the Spearman's correlation coefficient between the lower triangular parts of the neural and model RDM was computed (Fig. 3.1D) and then Fisher transformed.

Differences in activation patterns between days were compared using one-tailed paired t-test. We defined the ROIs as 16 mm radius spheres centered on MNI coordinates identified from previous literatures i.e., left pSTS (x, y, z = −52, −48, 2; Eisner et al. 2010), left IFG (–52, 16, 18; Eisner et al. 2010), and left vPM (-51, 0, 41; Evans and Davis 2015). To further reveal relationships between the neural representations in the frontotemporal regions, the representational connectivity analysis (RCA; Kriegeskorte et al. 2008b) was performed. We examined representational similarity between activation patterns in the ROIs by computing Spearman's correlation coefficient between neural RDMs in the ROIs (Fig. 3.1E).

To assess statistical differences in behavioral and fMRI data between days, the nonparametric bootstrap test, which does not require assumptions related to the distribution, used a pooled resampling method to account for small sample sizes (Dwivedi, Mallawaarachchi, & Alvarado, 2017). We thus performed a pooled bootstrap one-tailed paired t-test (10,000 bootstrap iterations).

3.3. Results

Behavioral performance improvement

The behavioral performance was scored as the proportion of correctly recognized mora of NVSS sentences (Fig. 3.2). A paired t-test showed that the mean performance across participants in the test on Day 7 was higher than in the first test on Day 1 [$t(4) = 5.97$, $p = 0.002$)]. In addition, the mean performance in the first test of the post-session was also higher than the performance in the first test on Day 1 [$t(4) = 4.75$, $p = 0.004$]. These results demonstrated that participants' listening performance of NVSS was improved

through the training and that the improvement was maintained in the post-session measurement.

Changes in cortical activity during listening to NVSS

The RSA analysis showed the representational content of brain regions by calculating a correlation between neural RDM, activation patterns to NVSS and clear speech, and the model RDM, distinguishing whether a stimulus was NVSS or clear speech (Fig. 3.1C). In order to test a change in the neural representations of NVSS across experimental days, we compared a correlation on Day 1 with a correlation on Day 7 (Fig. 3.3, top). We only found significant differences in the left pSTS ($t(4) = 2.49$, $p = 0.031$). In addition, to examine a slow and small change across days, we compared an average across correlations on days 1, 2, and 3 (early phase) with a correlation on days 5, 6, and 7 (late phase). A significant increase of the correlations in the late phase relative to the early phase were observed in the left pSTS ($t(4) = 6.00$, $p = 0.002$) and vPM ($t(4) = 8.48$, $p < 0.001$) (Fig. 3.3, top), indicating that the distinct activation patterns for NVSS from clear speech more clearly emerged in the late phase of experimental days than the early phase. The RCA analysis revealed that pattern similarity between neural RDMs of the left IFG and the left pSTS exhibited a significant difference between Day 1 and Day 7 ($t(4) = 2.89$, $p = 0.014$) (Fig. 3.3, bottom). Moreover, comparisons of the pattern similarities between the early phase and the late phase showed differences in the left pSTS and the left IFG pair ($t(4) = 2.64$, $p = 0.029$) and the left pSTS and the left vPM pair ($t(4) = 6.59$, $p = 0.002$) (Fig. 3.3, bottom). These results indicate that commonality in neural representation between the frontal and temporal regions was enhanced across experimental days.

Furthermore, we examined differences in correlations, between a neural RDM and the

model RDM, in the Day 1 session and post-session (Fig. 3.4). The neural activation patterns showed higher correlations in the post-session relative to the Day 1 session were found in the left IFG ($t(4) = 6.45$, $p = 0.002$) and the left pSTS ($t(4) = 2.27$, $p = 0.042$), suggesting that long-term effects of the neural changes for NVSS would be observed.

3.4. Discussion

In this study, we investigated changes in cortical activation underlying long-term perceptual learning of spectrally degraded speech. The behavioral results showed that participants improved their performance across experimental days, and the improvements persisted in the post-session (i.e., more than ten months later) (Fig. 3.2). We then provided a long-term characterization of continuous changes in neural representations for NVSS. The results of our RSA analysis showed that the neural activation patterns in the left frontotemporal regions for the NVSS condition compared to that for the clear speech condition changed across the experimental days (Fig. 3.3, top). Furthermore, the distinct activation patterns in the post-session were different from those in the first experiment day (Fig. 3.4).

The observed changes in behavioral performance and activity associated with perceptual learning across many days were consistent with changes seen in the visual modality survey (Aloufi et al., 2021; Bi et al., 2014; Chen et al., 2015, 2017; Frank et al., 2018, 2014; Yotsumoto et al., 2008). These activation changes have been observed in cortical regions involved in sensory and perceptual processing, such as the visual cortex (Aloufi et al., 2021; Chen et al., 2015; Frank et al., 2018, 2014; Yotsumoto et al., 2008), motion-sensitive middle temporal area (Chen et al., 2017), and the fusiform cortex (Bi et al., 2014). Even after a few weeks and years without training, changes in the behavioral

and cortical activations were maintained, indicating the maintenance of the reorganization of the sensory cortex playing a key role in enduring the improved perceptual performance (Aloufi et al., 2021; Bi et al., 2014; Chen et al., 2017; Frank et al., 2018, 2014). In line with these previous findings, the changes in activation patterns observed in the left pSTS, where speech-related sensory analysis is processed, may underlie the long-term maintenance of speech-perception improvement (Fig. 3.4).

Synaptic connection weights established via learning would be retained with continuous synaptic changes while learning task-irrelevant information in real life (Abraham & Robins, 2005). Thus, even after training ended, the once acquired neural representation to NVSS could be maintained for a long time.

Furthermore, in addition to the left pSTS, the left IFG exhibited that distinct activation pattern to NVSS in the post-session compared to the Day 1 session (Fig. 3.4). The activations in the left pSTS and the left IFG in the post-session may be influenced by reactivation of trained memory of perceptual processing. The previous visual perceptual learning study also showed the emergence of changes in the activation pattern in the high-level cortical region two weeks after training ended (Chen et al., 2017). It has been suggested that an additional task after training elicits reactivation of previously consolidated, trained memory and leads to further modification of the reactivated memory for reconsolidation (Censor, Sagi, & Cohen, 2012). Altmann and Young (1993) also suggested that persistence of speech perceptual performance resulted from the recall of trained perceptual processes.

The frontotemporal regions showed an increased commonality in neural representation with increases in experimental days (Fig. 3.3, bottom). Training degraded speech perception induced functional connectivity between the left pSTS and the left IFG

(Eisner et al., 2010) and similar temporal changes in activation of the left pSTS and the left vPM (Adank & Devlin, 2010). Sohoglu and Davis (2016) suggested that the interaction between the frontal and temporal regions underlying a predictive coding of the perception of degraded speech are proposed to support perceptual learning due to minimizing prediction error. In a visual perceptual learning study, a long-term training modulated the connectivity between the sensory and high-level cortical area, suggesting that the high-level cortical area refined the neural representation of trained stimuli in the sensory area (Chen et al., 2017). The observed changes in the representational connectivity between the left pSTS and the left IFG/vPM might reflect integration enhancements between the auditory sensory information and speech-related high-order information such as semantic context and articulation. On the other hand, while the activation pattern in the left pSTS and the left vPM ROIs changed in the later phase within the seven experimental days, the left IFG ROI did not show such a significant change and displayed a different shape of the RSA curve plot from the left vPM and the left pSTS (Fig. 3.3, top). The multiple time course of refining the local neural processing in degraded speech perception suggests that the processing of the frontotemporal regions was differentially recruited depending on a stage of long-term training.

These results demonstrated the potentials of training-induced optimization of cortical systems that assist the perception of degraded speech. It is necessary to further investigate their functional roles in perceptual learning processes; however, long-term neural measurements may provide us with which aspects of the perceptual processing have changed and when they have changed. Thus, it can be applied to objective characterizations of learning processes and the development of optimal rehabilitation programs.

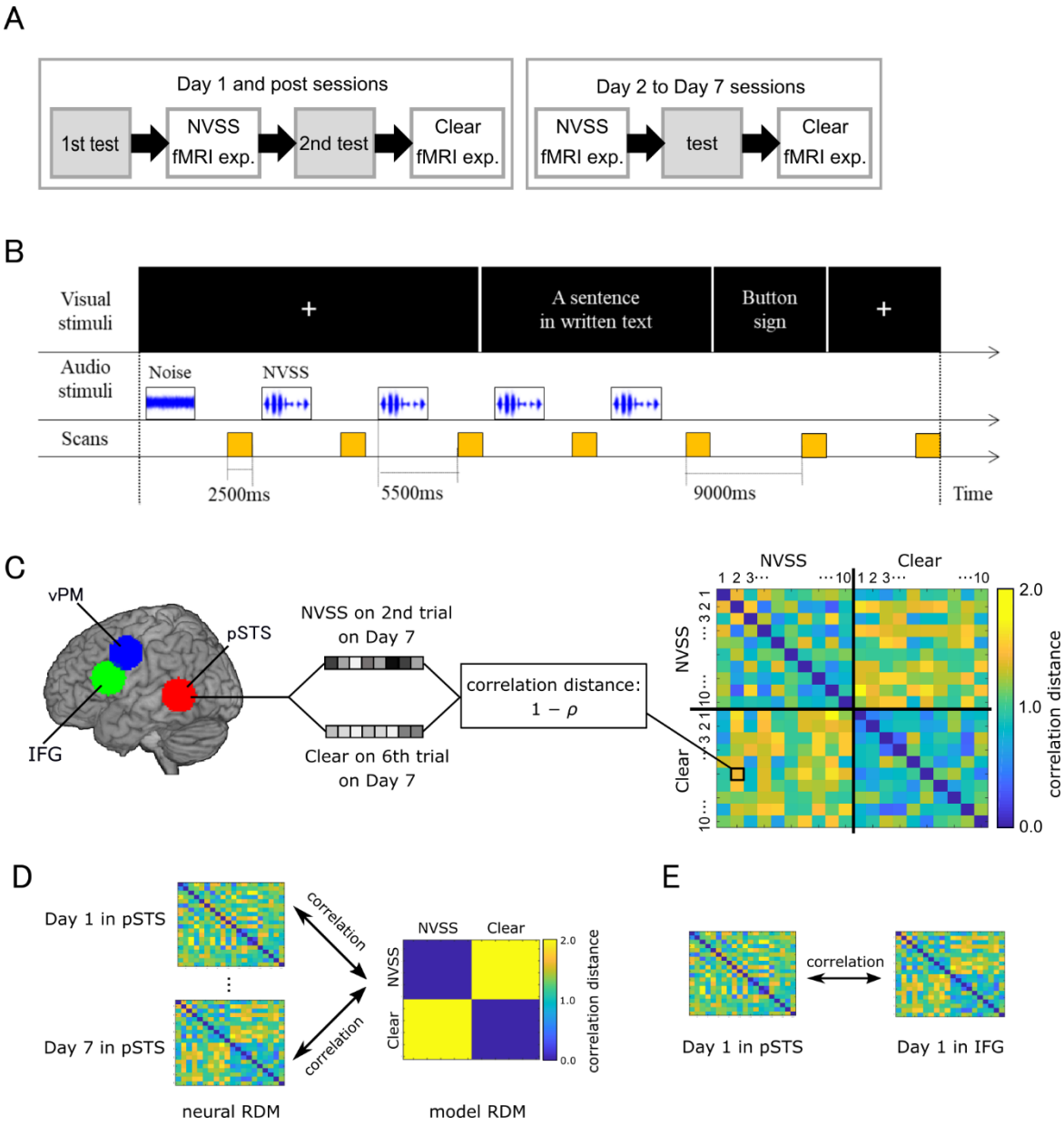

Fig. 3.1. (A) Experimental procedure. In the Day 1 session, participants underwent two behavioral tests and two fMRI experiments. From Day 2 to Day 7 sessions, one behavioral test and two fMRI experiments were conducted. In the post-session, which was performed after long periods (mean ± SD: 390 ± 78 days) of the end of the seven experimental days, two behavioral tests and two fMRI experiments were carried out. (B) Example of one trial in a NVSS session of an fMRI experiment. After notification sound (noise), NVSS was

repeated four times. In the first two NVSS presentations, participants were instructed to listen to NVSS carefully. In the next two presentations, participants were required to comprehend NVSS while viewing sentences on the screen. These processes allowed participants to efficiently learn perception of NVSS. (C–E) Procedure of the RSA and the RCA. fMRI activation patterns associated with each condition were abstracted from a ROI. A dissimilarity between the activation patterns associated with each condition is obtained as 1 minus the correlation. A neural RDM was then generated from dissimilarities for all pairs of conditions for each day (C). In RSA, correlation analyses were conducted between a neural RDM on each day and a model RDM representing stimulus types (i.e., NVSS and clear speech) to investigate how training of NVSS changed neural representations of speech perceptual processes (D). In RCA, neural RDMs across ROIs were compared for each day to reveal representational relationships between brain regions (E). vPM = left ventral premotor cortex; IFG = left inferior frontal gyrus; pSTS = left posterior superior temporal sulcus.

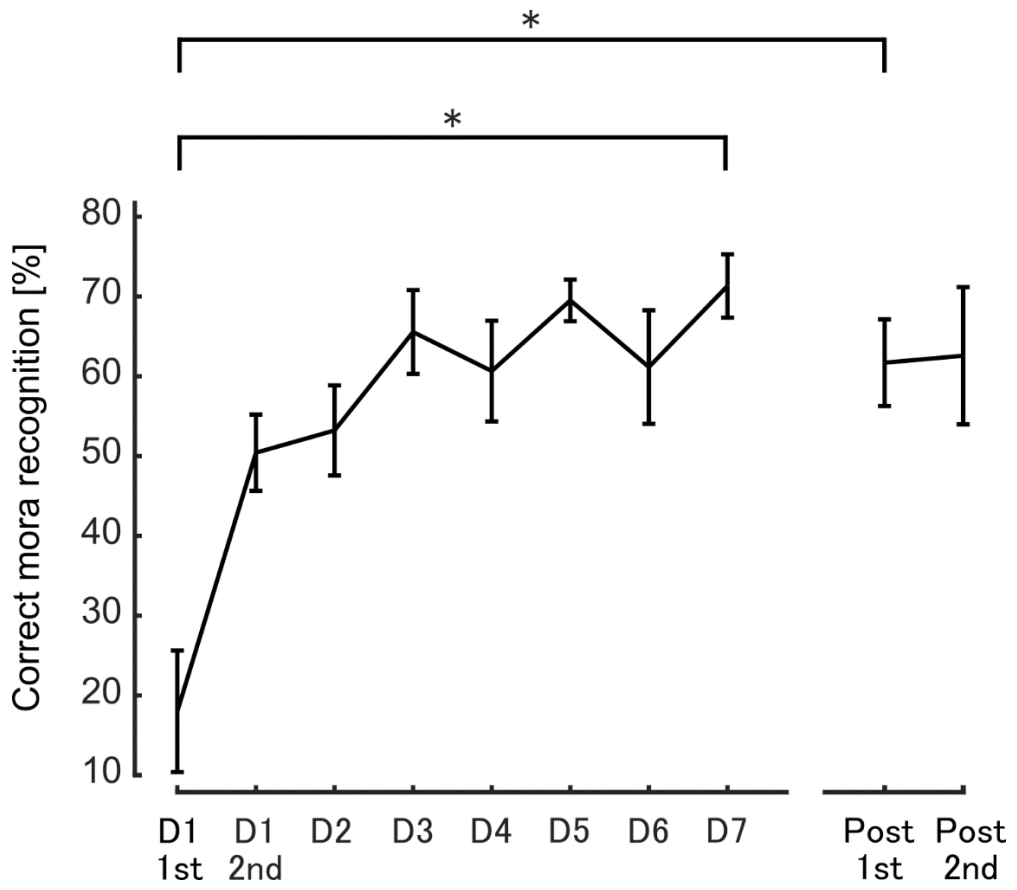

Fig. 3.2. Behavioral learning performance (correct mora recognition) across 10 tests. The mean performance across participants in the Day 7 session (mean = 71.3%, bootstrapped 95% confidence interval [CI] = [64.9%, 77.9%]) was significantly higher than the first test in the Day 1 session (mean = 18.0%, bootstrapped 95% CI = [7.12%, 32.5%]) (p = 0.002; one-sided paired t-test). The mean performance across participants in the first test in the post-session (mean = 61.7%, bootstrapped 95% CI = [51.7%, 70.8%]) was also higher than the performance in the first test in the Day 1 session (p = 0.005; one-sided paired t-test). Tests were conducted on the seven experimental days and in the post-session approximately one year later (mean ± SD: 390 ± 78 days). Error bars indicate 95% confidence intervals across participants, obtained by bootstrap (10,000 iterations).

The significant differences between sessions are displayed by asterisks (*$p < 0.05$; one-sided paired t-test). D1 = Day 1 session; Post = post-session.

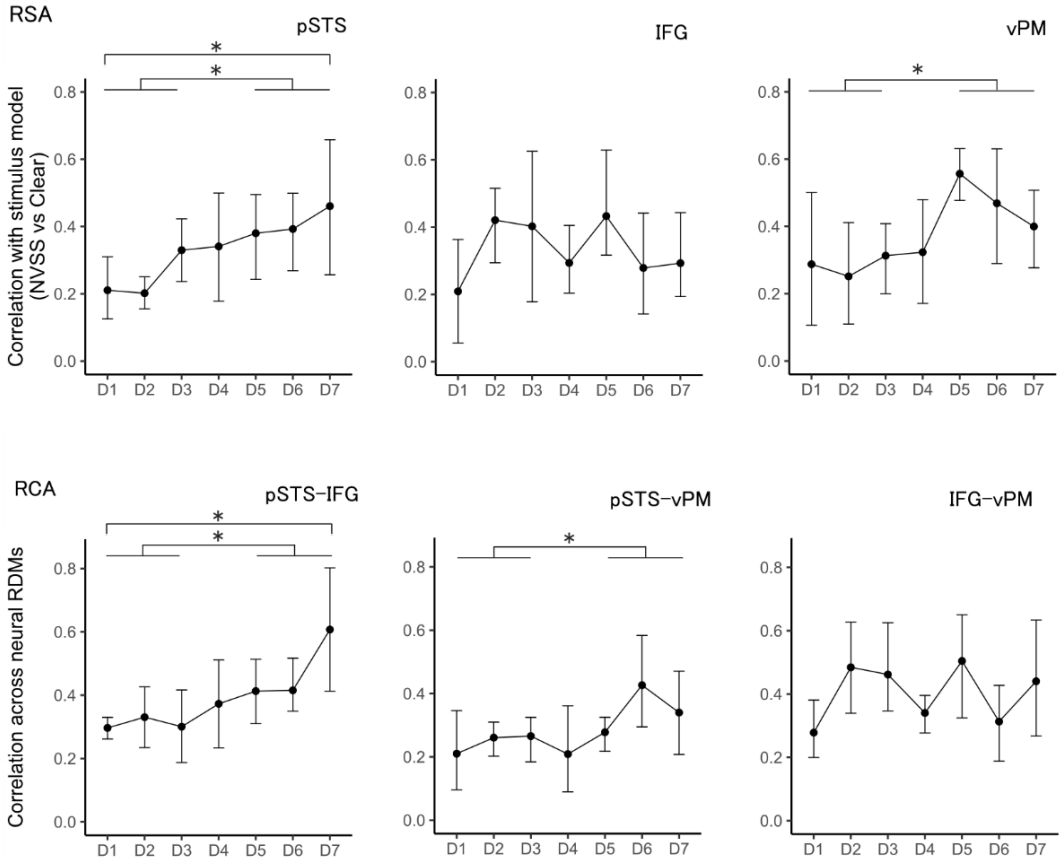

Fig. 3.3. Results of the representational similarity analysis (RSA, top) and the representational connectivity analysis (RCA, bottom) showing changes in fMRI activation patterns across seven experimental days. The left pSTS and the left vPM indicated that the differences in activation patterns for NVSS and clear speech got greater during experimental days (top). Representational connectivity between the left pSTS and the left IFG, and representational connectivity between the left pSTS and the left vPM increased across days (bottom). The significant differences between sessions are displayed by asterisks (*p < 0.05; one-sided paired t-test). Error bars indicate 95% confidence intervals, obtained by bootstrap (10,000 iterations). IFG = left inferior frontal gyrus; pSTS = left posterior part of superior temporal sulcus; vPM = left ventral premotor cortex. D1 = Day 1.

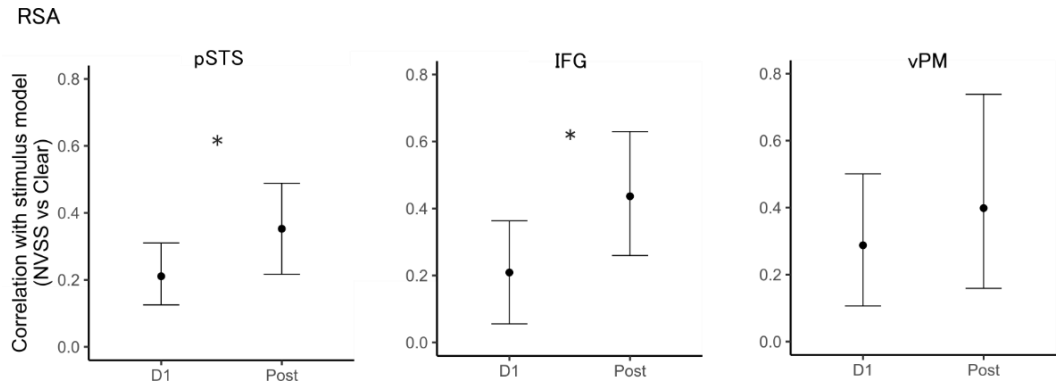

Fig. 3.4. Results of the representational similarity analysis (RSA) showing differences in fMRI activation patterns on Day 1 and in the post-session. The left pSTS and the left IFG indicated that the differences in activation patterns for NVSS and clear speech was more distinct in the post-session than Day 1. The significant differences between sessions are displayed by asterisks (*p < 0.05; one-sided paired t-test). Error bars indicate 95% confidence intervals, obtained by bootstrap (10,000 iterations). IFG = left inferior frontal gyrus; pSTS = left posterior part of superior temporal sulcus; vPM = left ventral premotor cortex. D1 = Day 1 session; Post = post-session.

4. General discussion

The current thesis investigated the neural mechanisms that underlie within-individual variability of the perception of acoustically degraded speech. By utilizing NVSS in normal-hearing participants, the behavioral and functional MRI studies were conducted to elucidate neural correlates of variability in the process of degraded speech perception within individuals. Furthermore, across-day perceptual training of NVSS demonstrated changes in behavioral performance and brain activity patterns, which would provide insights into cortical plasticity underlying perceptual processes in difficult listening situations and long-term rehabilitation of auditory disorders. These findings indicate that involvement of the frontotemporal cortices associated with the higher auditory and cognitive processes underpins variation in comprehension within individuals. This chapter first summarizes the main results of the experiments and discusses more specific findings in the dissertation.

4.1. Summary of the experimental results

In study 1, the relationships between fluctuations of NVSS comprehension within individuals and cortical responses were examined. The cerebral activity is associated with the degree of subjective comprehension of NVSS sentences using fMRI. Clear speech and noise trials were included as control conditions. The results showed greater activation in the right superior temporal cortex when participants recognized at least some words in an NVSS sentence. The left IFG exhibited increased activation when a listener recognized words in a sentence they did not fully comprehend. The laterality analysis further

indicated that less lateralized responses in the temporal cortex were observed to recognize words in an NVSS sentence; however, a left-lateralization was found when no words were recognized. These results revealed variability in neural activity of the frontal and temporal cortices associated with fluctuation in NVSS comprehension within individuals.

Study 2 investigated neural processes during and after long-term perceptual learning. The experiment conducted perceptual training of NVSS and measured the neural activation in the frontotemporal cortices associated with perceptual learning of acoustically degraded speech. Behavioral performance and fMRI activation patterns were collected for seven days and the postsession (more than ten months after the seven days) using fMRI. As a result, participants' performance of NVSS recognition improved across seven experimental days, and their performance in the postsession was also higher than the pretraining session. The representational similarity analysis revealed that the similarity between neural activation patterns to NVSS and clear speech on Day 7 got significantly different from Day 1 in the left pSTS. The left vPM in the representational similarity between NVSS and clear speech also exhibited changes across experimental days. Furthermore, the distinct neural activation patterns in the left pSTS were found when compared the postsession to Day 1. These results suggest that neural changes in the left pSTS were associated with improving and maintaining NVSS perception.

4.2. Variability of neural responses in the frontotemporal cortices for degraded speech perception

The interaction of the frontotemporal cortices is involved in higher-order processing for the perception of degraded speech (Obleser & Kotz, 2010; Sohoglu & Davis, 2016;

Sohoglu et al., 2012). It has been investigated based on external perceptual cues such as the presentation of clear speech or written text on the neural responses. Comparison between matching and mismatching cue conditions revealed the involvement of the frontotemporal cortices to degraded speech perception. Suppose the central nervous systems internally generate higher-order cues from their prior knowledge when listening to degraded speech. In that case, variations in the degree of matching between their internal cues as well as actual sensory inputs and that within-individual fluctuations in the frontotemporal interactions and NVSS comprehension occur even without external cues. The first study then assessed within-individual variability in the involvement of frontotemporal cortices for the process of NVSS perception. These changes in frontotemporal activity observed without external cues may be partially consistent with the frontotemporal interactions for the external cues in the previous studies and could be responsible for fluctuations in comprehension. In addition to trial-by-trial variation in the frontotemporal activity, the second study indicated that the relatively long-term neural changes in the frontotemporal cortices were induced through the experimental days conducting perceptual learning of degraded speech. The pattern similarity analysis showed changes in functional responses in the left pSTS, which is a region connecting between frontal and auditory cortices and relationships between pSTS and IFG as well as pSTS and vPM. Because the first study showed that, without perceptual training, the left IFG activity was elevated when a listener did not fully recognize NVSS, the left IFG could be involved in NVSS perception in the early stage of the training the second study. The involvement of the left IFG might continue for long-term training. Such continuous frontal processes might lead to modulation in neural processing in the left pSTS this study observed, following the idea of the predictive coding framework based on the comparison

between tow-down predictions of sensory input and the actual sensory input (Friston, 2005; Rao & Ballard, 1999; Sohoglu & Davis, 2016). Although further connectivity analyses are required to investigate the interactions between the frontal and auditory temporal cortices, these findings suggested that the frontotemporal interaction underlies the variation in comprehension and the process of perceptual learning of NVSS rather than local circuits within the auditory temporal cortices. This involvement of widespread neural resources might lead to rapid and long-lasting plastic changes in degraded speech perception.

In addition, these results demonstrated that within-individual conditions would lead to variable cortical activations during listening to NVSS. Therefore, future studies of degraded speech perceptual systems should take into account the control of acoustic properties related to the intelligibility of stimuli and visual stimuli presented as perceptual cues and changes in brain activity due to variation in comprehension and the process of perceptual learning within individuals.

4.3. Changes in neural representation of NVSS induced by perceptual training

The neural representations in the sensory and high-level cortical areas were changed through the perceptual training of NVSS in study 2. These changes were shown by the RSA using the RDM comparing NVSS and clear speech. Due to the changes in neural representation indicating the difference between NVSS, which are not fully intelligible but learnable sound, and clear speech, which is perfectly intelligible sound, if the changes in the neural representation were decreased via training days, the change would indicate linguistic processing similar to clear speech as byproducts of NVSS comprehension;

however, the changes were increased. Therefore, the changes would reflect new activation patterns to NVSS different from that to clear speech, suggesting that the changes in neural representation were learning-induced cortical plasticity. The physiological experiments in animal studies revealed that task-specific changes in neural responses were induced by training in auditory regions (e.g., Gao & Suga, 1998; Gentner & Margoliash, 2003; Polley, Steinberg, & Merzenich, 2006; Takahashi, Funamizu, Mitsumori, Kose, & Kanzaki, 2010). By performing electrophysiological recordings in rats, perceptual training enhanced the auditory cortex responses to task-relevant sounds characterized by frequency or intensity, and the degree of topographic map plasticity was correlated with the degree of perceptual learning (Polley et al., 2006). The left pSTS corresponds to Wernicke's area and is involved in processing familiar sounds and speech associated with meaning and articulation (Price, 2012). The studies of NVSS further indicated that the left pSTS is engaged in the process of syllable identification (Samuel Evans & Davis, 2015) and lexical information (Obleser & Kotz, 2010) and attention (Wild, Yusuf, et al., 2012) for the perception of degraded speech sounds. The perceptual training thus would induce the emergence of functional reorganization in the left pSTS to recognize degraded sounds as speech, although it is difficult to reveal details of the involvement of the pSTS from the current experiments.

4.4. Implications for the process of speech perception under various difficult conditions

Auditory experiments with normal-hearing participants using NVSS as an acoustic

simulation model of cochlear implants can offer insight into adaptation processes after implantation of the devices. While recipients of cochlear implants can take a long-term process of perceptual training after cochlear implantation, the training outcomes are limited and largely variable across children and adults with cochlear implants (Munson, Donaldson, Allen, Collison, & Nelson, 2003; Pisoni, 2000; Sarant, Blamey, Dowell, Clark, & Gibson, 2001). It is suggested that one of the factors influencing benefits that cochlear implant users receive is the plasticity of the brain (Moore & Shannon, 2009). Potential implications of the current research concern that neural measure of cortical responses associated with the degrees of speech comprehension and the training procedure could provide helpful information to monitor the perceptual process, suggesting biological markers of comprehension and learning progress. It is essential to design rehabilitation regimes to accelerate the speed and increase the beneficial outcome of the learning. However, these results may not be directly compared to the CI users' perception, because the current experiments only taking seven days of listening to NVSS was not as long as the actual training or adaptation of cochlear implant users. Therefore, the insights from experiments with acoustic simulation should be taken with caution.

Our findings of the perceptual processes of NVSS also have implications for shared mechanisms in speech perception under various acoustically degraded conditions. The previous study demonstrated that performance gains induced by training with NVSS could be generalized across different acoustic conditions (Hervais-Adelman, Davis, Johnsrude, Taylor, & Carlyon, 2011). The perceptual learning of bandpass filtered NVSS in one frequency range (50–1406 Hz or 1593–5000 Hz) transferred to an untrained frequency range indicated the generalizability of learning of NVSS across different frequency regions. The learning of vocoded speech with a carrier signal (noise bands,

pulse trains, or sine waves) partially transferred to the perception of vocoded speech with a different carrier signal indicated that improvements in perceptual process based on amplitude envelope cues could be generalized regardless of carrier signals (temporal fine structure). Moreover, performance on NVSS perception was correlated with time-compressed speech, which was a form of temporal distortion, and speech in babble noise, which was temporally and spectrally distorted (Carbonell, 2017). These results suggest that there are general abilities (e.g., working memory) that promote speech perception beyond spectral and temporal processing under various difficult listening conditions. Investigation of NVSS perception provides a link to such a shared mechanism, which might cover various types of auditory disorders and variability of inefficiencies in cochlear implant users due to signals provided by implant devices and device fitting, and residual hearing.

4.5. Limitation of the present research

There are a few limitations to this study. First, both studies were conducted using sentences stimuli, which contained redundant information as rich perceptual cues of speech sounds and did not dissociate between neural responses associated with the process of these cues. Speech perception of spoken sentences is associated with the processes of semantic and syntactic information of sentences and the processes of lexical and sublexical information of words contained in sentences. Therefore, it is necessary to examine neural processing of degraded speech perception by considering the effect of hierarchical structures in speech sounds, such as relationships between phonemic, syllabic, lexical, semantic, and syntactic information. This will further unveil functional roles of

the frontotemporal cortices for NVSS comprehension and learning.

In addition, further studies with large sample sizes are needed to assess these results from data with a small number of participants, which may negatively affect statistical power. The causal relationship between increased perceptual performance and neural changes was unclear from the results of this study. Therefore, future experiments of neural manipulation using transcranial magnetic stimulation (Hartwigsen, Golombek, & Obleser, 2015) and transcranial electric stimulation (tES) (Sehm et al., 2013), as well as collaboration with clinical studies, are required.

4.6. General conclusion

In this thesis, behavioral and neuroimaging measures to investigate the variability in neural activation associated with the perception of spectrally degraded speech processed with noise vocoder. The experiments demonstrated that degrees of subjective comprehension of a sentence fluctuate trial-by-trial when listening to NVSS with a single acoustic clarity (four-band noise vocoding) and neural activity in the speech network, including temporal and frontal regions differently correlates with the level of comprehension. In addition, an experiment employing long-term perceptual learning of NVSS across seven experimental days showed that listeners improved their performance, and the improvements persisted more than ten months later. The neural activation patterns in the left frontotemporal regions changed across the experimental days, and the change in the left pSTS activity was maintained for a long time, indicating a temporal characterization of cortical plasticity for NVSS. Taken together, these results suggest that the individual listening process of degraded speech sounds is underpinned by the perceptual processing in the frontotemporal regions, as reflected by cortical lateralization and changes in neural representation.

Understanding of perceptual mechanisms of NVSS could indicate better recognition processes for degraded auditory inputs, suggesting appropriate use of neural measurements to observe processes during speech perception (Lawrence, Wiggins, Anderson, Davies-thompson, & Hartley, 2018; Wijayasiri, Hartley, & Wiggins, 2017). In addition, further studies taking into account within-individual variability would help characterize the central nervous system for the process of speech signals, shedding light on rehabilitation of patients with cochlear implants, hearing aids, or other listening difficulties.

References

Abraham, W. C., & Robins, A. (2005). Memory retention - The synaptic stability versus plasticity dilemma. *Trends in Neurosciences*, *28*(2), 73–78. https://doi.org/10.1016/j.tins.2004.12.003

Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage*, *49*(1), 1124–1132. https://doi.org/10.1016/j.neuroimage.2009.07.032

Aloufi, A. E., Rowe, F. J., & Meyer, G. F. (2021). Behavioural performance improvement in visuomotor learning correlates with functional and microstructural brain changes. *NeuroImage*, *227*(March 2020), 117673. https://doi.org/10.1016/j.neuroimage.2020.117673

Altmann, G. T. M., & Young, D. (1993). Factors Affecting Adaptation to Time-Compressed Speech. *Speech Communication*, (September), 1359–1362.

Bi, T., Chen, J., Zhou, T., He, Y., & Fang, F. (2014). Function and structure of human left fusiform cortex are closely associated with perceptual learning of faces. *Current Biology*, *24*(2), 222–227. https://doi.org/10.1016/j.cub.2013.12.028

Blank, H., & Davis, M. H. (2016). Prediction errors but not sharpened signals simulate multivoxel fMRI patterns during speech perception. *PLoS Biology*, *14*(11), e1002577. https://doi.org/10.1371/journal.pbio.1002577

Bradshaw, A. R., Bishop, D. V. M., & Woodhead, Z. V. J. (2017). Methodological considerations in assessment of language lateralisation with fMRI: A systematic review. *PeerJ*, *2017*(7). https://doi.org/10.7717/peerj.3557

Carbonell, K. M. (2017). Reliability of individual differences in degraded speech

perception. *The Journal of the Acoustical Society of America*, *142*(5), EL461–EL466. https://doi.org/10.1121/1.5010148

Censor, N., Sagi, D., & Cohen, L. G. (2012). Common mechanisms of human perceptual and motor learning. *Nature Reviews Neuroscience*, *13*(9), 658–664. https://doi.org/10.1038/nrn3315

Chen, N., Bi, T., Zhou, T., Li, S., Liu, Z., & Fang, F. (2015). Sharpened cortical tuning and enhanced cortico-cortical communication contribute to the long-term neural mechanisms of visual motion perceptual learning. *NeuroImage*, *115*, 17–29. https://doi.org/10.1016/j.neuroimage.2015.04.041

Chen, N., Lu, J., Shao, H., Weng, X., & Fang, F. (2017). Neural mechanisms of motion perceptual learning in noise. *Human Brain Mapping*, *38*(12), 6029–6042. https://doi.org/10.1002/hbm.23808

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *The Journal of Neuroscience*, *23*(8), 3423–3431. https://doi.org/10.1523/JNEUROSCI.23-08-03423.2003

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology. General*, *134*(2), 222–241. https://doi.org/10.1037/0096-3445.134.2.222

Dorman, M. F., Loizou, P. C., & Rainey, D. (1997). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *The Journal of the Acoustical Society of America*, *102*(5), 2993–2996. https://doi.org/10.1121/1.420354

Dwivedi, A. K., Mallawaarachchi, I., & Alvarado, L. A. (2017). Analysis of small sample size studies using nonparametric bootstrap test with pooled resampling method. *Statistics in Medicine*, *36*(14), 2187–2205. https://doi.org/10.1002/sim.7263

Eckert, M. A., Teubner-Rhodes, S., & Vaden, K. I. (2016). Is Listening in Noise Worth It? The Neurobiology of Speech Recognition in Challenging Listening Conditions. *Ear and Hearing*, *37*(11), 101S-110S. https://doi.org/10.1097/AUD.0000000000000300

Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., & 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. https://doi.org/10.1523/JNEUROSCI.4040-09.2010

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, *113*(28), 7900–7905. https://doi.org/10.1073/pnas.1602413113

Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *Journal of Neuroscience*, *33*(26), 10688–10697. https://doi.org/10.1523/JNEUROSCI.4596-12.2013

Erb, J., & Obleser, J. (2013). Upregulation of cognitive control networks in older adults' speech comprehension. *Frontiers in Systems Neuroscience*, *7*, 116. https://doi.org/10.3389/fnsys.2013.00116

Evans, S., Kyong, J. S., Rosen, S., Golestani, N., Warren, J. E., McGettigan, C., … Scott, S. K. (2014). The pathways for intelligible speech: Multivariate and

univariate perspectives. *Cerebral Cortex*, *24*(9), 2350–2361. https://doi.org/10.1093/cercor/bht083

Evans, Samuel, & Davis, M. H. (2015). Hierarchical Organization of Auditory and Motor Representations in Speech Perception: Evidence from Searchlight Similarity Analysis. *Cerebral Cortex*, *25*(12), 4772–4788. https://doi.org/10.1093/cercor/bhv136

Fairbanks, G., & Kodman, F. (1957). Word Intelligibility as a Function of Time Compression. *The Journal of the Acoustical Society of America*, *29*(5), 636–641. https://doi.org/10.1121/1.1908992

Frank, S. M., Greenlee, M. W., & Tse, P. U. (2018). Long Time No See: Enduring Behavioral and Neuronal Changes in Perceptual Learning of Motion Trajectories 3 Years After Training. *Cerebral Cortex (New York, N.Y. : 1991)*, *28*(4), 1260–1271. https://doi.org/10.1093/cercor/bhx039

Frank, S. M., Reavis, E. A., Tse, P. U., & Greenlee, M. W. (2014). Neural mechanisms of feature conjunction learning: Enduring changes in occipital cortex after a week of training. *Human Brain Mapping*, *35*(4), 1201–1211. https://doi.org/10.1002/hbm.22245

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Fu, Q.-J., Zeng, F.-G., Shannon, R. V, & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *The Journal of the Acoustical Society of America*, *104*(1), 505–510. https://doi.org/10.1121/1.423251

Gao, E., & Suga, N. (1998). Experience-dependent corticofugal adjustment of midbrain

frequency map in bat auditory system. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(21), 12663–12670. https://doi.org/10.1073/pnas.95.21.12663

Gentner, T. Q., & Margoliash, D. (2003). Neuronal populations and single cells representing learned auditory objects. *Nature*, *424*(6949), 669–674. https://doi.org/10.1038/nature01731

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., … Bowtell, R. W. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping*, *7*(3), 213–223. https://doi.org/10.1002/(SICI)1097-0193(1999)7:3<213::AID-HBM5>3.0.CO;2-N

Hartwigsen, G., Golombek, T., & Obleser, J. (2015). Repetitive transcranial magnetic stimulation over left angular gyrus modulates the predictability gain in degraded speech comprehension. *Cortex*, *68*, 100–110. https://doi.org/10.1016/j.cortex.2014.08.027

Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I. S., & Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Language and Cognitive Processes*, *27*(7–8), 1145–1166. https://doi.org/10.1080/01690965.2012.662280

Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: effects of feedback and lexicality. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(2), 460–474. https://doi.org/10.1037/0096-1523.34.2.460

Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., Taylor, K. J., & Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *Journal of*

Experimental Psychology: Human Perception and Performance, *37*(1), 283–295. https://doi.org/10.1037/a0020772

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Karni, a, & Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(11), 4966–4970. https://doi.org/DOI 10.1073/pnas.88.11.4966

Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(4), 1–28. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., … Bandettini, P. A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

Kyong, J. S., Scott, S. K., Rosen, S., Howe, T. B., Agnew, Z. K., & McGettigan, C. (2014). Exploring the roles of spectral detail and intonation contour in speech intelligibility: an fMRI study. *Journal of Cognitive Neuroscience*, *26*(8), 1748–1763. https://doi.org/10.1162/jocn_a_00583

Lawrence, R. J., Wiggins, I. M., Anderson, C. A., Davies-thompson, J., & Hartley, D. E. H. (2018). Cortical correlates of speech intelligibility measured using functional near-infrared spectroscopy (fNIRS). *Hearing Research*, *370*, 53–64. https://doi.org/10.1016/j.heares.2018.09.005

Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & Cramon, D. Y. Von. (2002).

FMRI Reveals Brain Regions Mediating Slow Prosodic Modulations in Spoken Sentences, *88*, 73–88. https://doi.org/10.1002/hbm.10042

Moore, D. R., & Shannon, R. V. (2009). Beyond cochlear implants: Awakening the deafened brain. *Nature Neuroscience*, *12*(6), 686–691. https://doi.org/10.1038/nn.2326

Munson, B., Donaldson, G. S., Allen, S. L., Collison, E. A., & Nelson, D. A. (2003). Patterns of phoneme perception errors by listeners with cochlear implants as a function of overall speech perception ability. *The Journal of the Acoustical Society of America*, *113*(2), 925–935. https://doi.org/10.1121/1.1536630

Nakajima, Y., Matsuda, M., Ueda, K., & Remijn, G. B. (2018). Temporal Resolution Needed for Auditory Communication: Measurement With Mosaic Speech. *Frontiers in Human Neuroscience*, *12*(April), 1–8. https://doi.org/10.3389/fnhum.2018.00149

Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, *20*(3), 633–640. https://doi.org/10.1093/cercor/bhp128

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., … Hickok, G. (2010). Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, *20*(10), 2486–2495. https://doi.org/10.1093/cercor/bhp318

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, *10*(27), 1–27. https://doi.org/10.3389/fninf.2016.00027

Pisoni, D. B. (2000). Cognitive factors and cochlear implants: Some thoughts on perception, learning, and memory in speech perception. *Ear and Hearing*, *21*(1), 70–78. https://doi.org/10.1097/00003446-200002000-00010

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time." *Speech Communication*, *41*(1), 245–255. https://doi.org/10.1016/S0167-6393(02)00107-3

Polley, D. B., Steinberg, E. E., & Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *Journal of Neuroscience*, *26*(18), 4970–4982. https://doi.org/10.1523/JNEUROSCI.3771-05.2006

Price, C. J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Remez, R., Rubin, P., Pisoni, D., & Carrell, T. (1981). Speech perception without traditional speech cues. *Science*, *212*(4497), 947–949. https://doi.org/10.1126/science.7233191

Riquimaroux, H. (2006). Perception of noise-vocoded speech sounds: Sentences, words, accents and melodies. *Acoustical Science and Technology*, *27*(6), 325–331. https://doi.org/10.1250/ast.27.325

Saberi, K., & Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature*, *398*(6730), 760. https://doi.org/10.1038/19652

Sarant, J. Z., Blamey, P. J., Dowell, R. C., Clark, G. M., & Gibson, W. P. R. (2001). Variation In Speech Perception Scores Among Children with Cochlear Implants. *Ear and Hearing*, *22*(1), 18–28. https://doi.org/10.1097/00003446-200102000-00003

Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some Effects of Training on the Preception of Synthetic Speech. *Human Factors*, *27*(4), 395–408. https://doi.org/10.1177/001872088502700404

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(12), 2400–2406. https://doi.org/10.1093/brain/123.12.2400

Scott, S. K., & Johnsrude, I. S. (2003, March). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*. https://doi.org/10.1016/S0166-2236(02)00037-1

Sehm, B., Schnitzler, T., Obleser, J., Groba, A., Ragert, P., Villringer, A., & Obrig, H. (2013). Facilitation of inferior frontal cortex by transcranial direct current stimulation induces perceptual learning of severely degraded speech. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(40), 15868–15878. https://doi.org/10.1523/JNEUROSCI.5466-12.2013

Shannon, Robert V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303–304. https://doi.org/10.1126/science.270.5234.303

Shannon, R V, Zeng, F. G., & Wygonski, J. (1998). Speech recognition with altered spectral distribution of envelope cues. *The Journal of the Acoustical Society of America*, *104*(4), 2467–2476. https://doi.org/10.1121/1.423774

Smalt, C. J., Gonzalez-Castillo, J., Talavage, T. M., Pisoni, D. B., & Svirsky, M. A. (2013). Neural correlates of adaptation in freely-moving normal hearing subjects under cochlear implant acoustic simulations. *NeuroImage, 82*, 500–509. https://doi.org/10.1016/j.neuroimage.2013.06.001

Sohoglu, E., & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences*, *113*(12), E1747–E1756. https://doi.org/10.1073/pnas.1523266113

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive Top-Down Integration of Prior Knowledge during Speech Perception. *Journal of Neuroscience*, *32*(25), 8443–8453. https://doi.org/10.1523/JNEUROSCI.5069-11.2012

Tachibana, R. O., Sasaki, Y., & Riquimaroux, H. (2013). Relative contributions of spectral and temporal resolutions to the perception of syllables, words, and sentences in noise-vocoded speech. *Acoustical Science and Technology*, *34*(4), 263–270. https://doi.org/10.1250/ast.34.263

Takahashi, H., Funamizu, A., Mitsumori, Y., Kose, H., & Kanzaki, R. (2010). Progressive plasticity of auditory cortex during appetitive operant conditioning. *BioSystems, 101*(1), 37–41. https://doi.org/10.1016/j.biosystems.2010.04.003

Ueda, K., Araki, T., & Nakajima, Y. (2018). Frequency specificity of amplitude envelope patterns in noise-vocoded speech. *Hearing Research, 367,* 169–181. https://doi.org/10.1016/j.heares.2018.06.005

Ueda, K., & Nakajima, Y. (2017). An acoustic key to eight languages/dialects: Factor analyses of critical-band-filtered speech. *Scientific Reports, 7*, 1–4. https://doi.org/10.1038/srep42468

Vigneau, M., Beaucousin, V., Hervé, P. Y., Jobard, G., Petit, L., Crivello, F., … Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage*, *54*(1), 577–593. https://doi.org/10.1016/j.neuroimage.2010.07.036

Wijayasiri, P., Hartley, D. E. H., & Wiggins, I. M. (2017). Brain activity underlying the recovery of meaning from degraded speech: A functional near-infrared spectroscopy (fNIRS) study. *Hearing Research*, *351*, 55–67. https://doi.org/10.1016/j.heares.2017.05.010

Wild, C. J., Davis, M. H., & Johnsrude, I. S. (2012). Human auditory cortex is sensitive to the perceived clarity of speech. *NeuroImage*, *60*(2), 1490–1502. https://doi.org/10.1016/j.neuroimage.2012.01.035

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful Listening: The Processing of Degraded Speech Depends Critically on Attention. *Journal of Neuroscience*, *32*(40), 14010–14021. https://doi.org/10.1523/JNEUROSCI.1528-12.2012

Wilke, M., & Lidzba, K. (2007). LI-tool: A new toolbox to assess lateralization in functional MR-data. *Journal of Neuroscience Methods*, *163*(1), 128–136. https://doi.org/10.1016/j.jneumeth.2007.01.026

Xu, L., Thompson, C. S., & Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *The Journal of the Acoustical Society of America*, *117*(5), 3255–3267. https://doi.org/10.1121/1.1886405

Xu, L., Tsai, Y., & Pfingst, B. E. (2002). Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *The Journal of the Acoustical*

Society of America, *112*(1), 247. https://doi.org/10.1121/1.1487843

Yotsumoto, Y., Watanabe, T., & Sasaki, Y. (2008). Different Dynamics of Performance and Brain Activation in the Time Course of Perceptual Learning. *Neuron*, *57*(6), 827–833. https://doi.org/10.1016/j.neuron.2008.02.034

Curriculum vitae

Shota Murai

Education:

M.Sc., Engineering, Doshisha University, Kyoto, Japan, 2015

B.Sc, Engineering, Doshisha University, Kyoto, Japan, 2013

Research experience

Research Assistant, Doshisha University, 2018–2019

JSPS Research Fellowship for Young Scientists DC1, Japan Society for the Promotion of Science, 2015 – 2018

Awards:

The 2021 IEEE 3rd Global Conference on Life Sciences and Technologies Silver Prize, 2021

Most Outstanding Poster Award, the General Meeting of Evolinguistics, 2019

Outstanding Presentation Award, the Annual Meeting of the Society for Bioacoustics, 2014

Poster Award, National Institute for Physiological Sciences Research Meeting, 2013

Grants and Fellowships:

Grant-in-Aid for JSPS Fellows DC1, 2015–2018