이학석사 학위논문

Decoding Higher-level Language Processing During Semantic Speech Perception

음성 의미 지각시의 고등 언어 성분 처리 디코딩

2022년 8월

서울대학교 대학원
협동과정 뇌과학전공

김 의 영

Decoding Higher-level Language Processing During Semantic Speech Perception

지도교수 정 천 기

이 논문을 이학석사 학위논문으로 제출함
2022년 8월

서울대학교 대학원
협동과정 뇌과학전공

김 의 영

김의영의 석사 학위논문을 인준함
2022년 7월

위 원 장 ______김 남 수______ (인)

부위원장 ______정 천 기______ (인)

위 원 ______이 성 은______ (인)

Abstract

Decoding Higher-level Language Processing During Semantic Speech Perception

EuiYoung, Kim
Interdisciplinary Program in Neuroscience
The Graduate School
Seoul National University

High-level linguistic processing in the human brain remains incompletely understood and constitutes a challenging topic in speech neuroscience. While most studies focused on decoding low-level phonetic components using intracranial recordings of the human brain during speech perception, few studies have attempted to decode high-level syntactic or semantic features. If any, most of the research targeting semantic decoding is conducted with picture naming tasks, which only deal with visual language rather than spoken language.

The presenting study is focused on better characterizing the neural representations of processing spoken language perception, namely speech perception. Especially not on the lower-level language components such as phonemes or phonetics, but the higher-level components such as syntax and semantics. Since it is widely accepted that the tripartite nature of language processing consists of phonology, syntax, and semantics, a strategical method for analyzing speech perception tasks that can reject the intervention of phonetic factors was mandatory. Therefore, we conducted a question-and-answer speech task containing four questions revolving around two semantic categories (alive, body parts) with phonetically controlled words.

Intracranial neural signals were recorded during the question-and-answer speech task using electrocorticography (ECoG) electrodes for 14 epilepsy patients. Post hoc brain activity analysis was conducted for three subjects who answered correctly to every trial (144 trials in total) to ensure the analyzed data contained only brain signals collected during the correct semantic processing. The decoding results suggest that absolute and relative spectral neural feature trends occur across all participants in particular time windows. Furthermore, the spatial aspect of the neural features that yield the best decoding accuracy verifies the current biophysiological brain language model explaining the circular nature of word meaning comprehension in the left hemisphere language network.

Keyword: Human brain, Semantic, Speech perception, Electrocorticogram, Decoding, Classification

Student Number: 2020-22887

Contents

1. Introduction

1.1. Study Background

Among various complex cognitive functions of human brains, how the brain processes spoken language remains one of the most challenging topics in the field of brain decoding. The traditional Wernicke-Lichtheim-Geschwind language processing model involves Broca's area in the inferior frontal gyrus for language production and Wernicke's area in the posterior superior temporal gyrus for language comprehension. The two core language areas exchange information via the arcuate fasciculus, a bundle of fiber connecting frontal (Broca's area) and temporal (Wernicke's area) cortices[12, 14].

More recent theories integrate multiple brain cortices for more complex functions during language processing. For example, the Memory, Unification and Control (MUC) model suggests the involvement of the temporal and parietal cortex for the retrieval of building blocks of language, such as syntactic (lemma) templates and semantic (conceptual) meanings[14]. Following the information retrieval, further unification of holistic language processing occurs in the inferior frontal cortex areas, especially near the Broca's area. Various conventional functional magnetic resonance imaging (fMRI) lesion and experimental studies have shown that such intricate cooperation of multiple brain areas is indeed happening at a swift pace of a few hundred milliseconds. Some of the most typical language-related brain responses captured in event-related potentials (ERPs) are early left anterior negativity (ELAN), N400, and P600, and they are believed to be closely associated with syntactic, semantic, and integrative processes, respectively. Onsets of each ERPs listed above are roughly 100-150ms (ELAN), 250ms (N400), and 500ms (P600) after the acoustic word onset[12, 14]. This suggests that higher-level language factors (syntactic, semantic, and unification) are actively processed well before the word offset. Likewise, several speech perception fMRI studies have successfully established a broad spatiotemporal understanding of when and where the brain builds up syntactic frames and retrieves lemma to conceive a semantic concept[7,8,12,17-21,].

In recent years, several leading brain-computer interface (BCI) research groups have attempted to decode speech perception and production using intracranial electroencephalography (iEEG) electrodes such as electrocorticography (ECoG) grids and stereo-electroencephalography (SEEG) depths[1-6,16,24-28,32-34,36]. While all these iEEG studies successfully captured phonetic aspects of speech perception, few attempts have been made regarding the semantic speech decoding with invasive intracranial electrodes. Most of the studies on decoding linguistic semantics using iEEG were performed with picture naming tasks, focusing on visual language semantics[37,42]. One recent study has shown that frequency-dependent cortical interactions between the semantic hub at the ventral anterior temporal lobe (vATL) are associated with semantic processing during a picture naming task42. However, it remains unclear whether a semantic hub is required to process speech perception or semantic representations are generated from neural signals across the multi-dimensional space.

1.2. Purpose of Research

In the present study, we attempt to enhance the understanding of semantic contexts being processed in natural speech perception by identifying key neural features related to semantic language processing during a question-and-answer task. Neural activities are recorded while the subject actively engages in the task where they must proactively process the question they listen to. Then by classifying neural signals recorded during the phonetically controlled part of the question, we could verify the optimal neural feature that highly engages in discerning the two semantically opposing concepts. The traits of neural features we desire to identify in the study include spectral, temporal, and spatial information. Utilizing the high spatiotemporal resolution of ECoG electrodes, the neural signals acquired during the task will be inspected in terms of milliseconds to verify the current speech perception models thoroughly.

2. Materials and Methods

2.1. Human Subjects

This study was approved by the Institutional Review Board of Seoul National University Hospital (H-2011-087-1173) and was conducted after obtaining informed consent from all the subjects. Fourteen subjects participated in this study and were undergoing ECoG monitoring for clinical purposes of localizing the epileptogenic zone (Table 2). Intracranial grids, strips, and depth electrodes were placed across brain cortices, covering core brain regions related to language processing, such as the inferior frontal lobe and perisylvian region. Precise electrode coverage differs for each patient for clinical reasons. Detailed information about each patient is presented in table S1. Electrode localization was determined by CT-MRI co-registration.

After data collection, particular subjects suitable for semantic analysis were selected. For the sake of training the semantic decoder model with as many trials as possible, subjects who answered the semantic trials correctly without any wrong trials were selected for the analysis in this study. That is because answering wrong answers refers to having an erroneous semantic perception of the question given in the form of speech. The selected patients for the semantic speech perception analysis are patients number 8, 10, and 14.

2.2. Experimental design for ECoG data acquisition

The task was designed to ensure the subjects' active engagement in semantic speech understanding by establishing a context of being asked a question. Compared to being passively exposed to speech sentences, being an agent participating in a question-answer conversational context will ensure the subjects' proactive semantic understanding of the speech stimuli. In the task, the subject listened to one of four different questions using a loudspeaker while nothing was shown on the screen, looked at the two icons on a screen, and answered the questions by saying the name of the correct icon (Figure 1). The four separate questions (Q1-4) asked which of the following two icons is Q1: alive ('+ life'), Q2: not alive ('- life'), Q3: a part of a body ('+ body'), and Q4: not a part of a body ('- body') (Q1: "다음 중 살아있는 것은 무엇입니까?", Q2: "다음 중 살아있지 않은 것은 무엇입니까?", Q3: "다음 중 몸의 일부인 것은 무엇입니까?", Q4; "다음 중 몸의 일부가 아닌 것은 무엇입니까?"). The semantic experiment consists of 8 sessions, and each consists of 18 trials (total of 8*18 = 144 trials per participant). The experiment is class-balanced in that it consists of all four questions equally, with 36 trials per question. Throughout the trials, four different questions are presented to the subject in a pseudo-random order. Each trial consists of four phases: 3s-rest (gray fixation cross), 5s-listen (gray fixation cross + question played by speakers), 2s-think (two icons displayed), and 3s-answer (green fixation cross). The study aims to identify neural features that best classify two semantic categories: 'life' (Q1 vs. Q2, or +life vs. -life) and ‘body part’ (Q3 vs. Q4, or +bodv vs. -body).

2.3. Feature selection

To classify +life vs. -life and +body vs. -body speech perception neural signals based on their higher language components, such as syntactic and semantic neural features, the primary factor to consider is to control the phonetic components. Therefore, all question sentences were broken down into morphemes and were selected for the phonetically identical parts. That is, across questions 1-4, the phonetically identical segments that contain sustained core semantic meaning are ‘것은 (is)’, and ‘무엇입니까? (which)’. For Q1 and Q2, each ‘것은 (is)’, and ‘무엇입니까? (which)’ should contain the sustained meaning of ‘alive’ and ‘not alive’. The same applies to Q3 and Q4 as well, with different semantic meanings of ‘is a body part’ and ‘is not a body part’. Neural signals obtained during the listed two word segments for each question were bipolar re-referenced and band-pass filtered into seven different frequency bands (delta: 0.5-4Hz, theta: 4-8Hz, alpha: 8-12Hz, beta: 12-30Hz, gamma: 30-50Hz, high-gamma 1: 70-110Hz, and high-gamma 2: 130-170Hz). Then each filtered signal was calculated for power spectral density (PSD) by computing the rolling variance with a time window of 100ms. Finally, each PSD value was segmented with a 50ms time window interval for each word segment (Number of time windows: for +life vs. -life, ‘것은 (is)’: 19, ‘무엇입니까? (which)’: 39, and for +body vs. -body, ‘것은 (is)’: 18, ‘무엇입니까? (which)’: 38). PSD values were then averaged across each time window. Finally, for the number of each subject's electrode channels after bad channel rejection (N), combination of 2 channels (NC2) were generated and the two selected combination of neural features (PSD) were together used as the inputs for the linear discriminant analysis (LDA) classifier discriminating the two semantic classes (+life vs. -life or +body vs. -body). Classification accuracies were validated with 4-fold cross-validation. Computational analysis was achieved using Python (Python 3.9).

2.4. Apparatus

Intracranial EEG data were recorded at 2kHz using Neuvo amplifiers and the ProFusion EEG5 software (Compumedics, Australia). Speech audio was acquired at 48kHz using the built-in "Voice Memo" application from an iPhone placed close to the study participants. StimTracker (Cedrus, USA) sent identical time markers to both audio and brain signal recordings. The time markers were used in post hoc analysis to sync the two types of recordings.

3. Results

3.1. Visualization of spectral dynamics during the speech task

Time-frequency analysis of brain oscillations can be achieved by applying the continuous wavelet transform. To verify if the designed speech task does induce desired spectral dynamics, trial-averaged neural data were plotted for subject 7 (Figure 2). Subject 7 was chosen for time-frequency analysis because the subject had a long 1x8 ECoG strip that spans across the whole posterior temporal lobe with parts of the occipital lobe and parietal lobe as well. The figures display distinctive task-related neural activity tendencies. Electrodes 11 and 12, located in the occipital lobe, show active gamma-high gamma activation during the icon presentation. Electrodes 13 to 15, located in the inferior to middle temporal gyrus, are generally more delta and alpha activity oriented during the speech task. Interestingly, electrodes 16 and 17, located in the superior temporal gyrus and Heschel's gyrus each, display an outstanding gamma-high gamma activity during speech perception and production. The significant difference between the electrode 16 and 17 is that in electrode 17, there is a distinctive deactivation in the upper theta and activation in the delta range. Visualized neural activities during the task confirm the dynamic engagement of each frequency band during the designed question-answer speech task.

3.2. Semantic classification accuracy for speech perception

Figure 3 depicts each patient's highest accuracy for discerning two semantic categories: +life vs. -life and +body vs. -body. The chance level accuracy for both semantic categories is 50% since both are binary classification. At the top of each bar graph are summaries of the neural feature that yielded the best classification performance in spectral, spatial, and temporal aspects. For example, figure 3 illustrates that subject 8 recorded the highest decoding accuracy (79.2%) by far for classifying +life vs. -life semantic category when using the neural signals during word segment' 무엇입니까? (Which)', alpha frequency band, from two different locations of the amygdala (AMG) and inferior frontal gyrus (IFG), during the 40th time window (2000ms-2050ms) as an input to the classifier.

Another way to analyze the classification accuracy trend is by plotting the accuracy change in the time order (Figure 4). Assuming that the conventional theory of speech perception is correct, the graph backgrounds are marked in pink, blue, and purple according to the onset of syntactic (100-150ms), semantic (250-500ms), and unification (500-750ms) processing, respectively. For example, in figure 4A, subject 8 has several accuracy peaks characteristic of each syntactic, semantic, and unification processing epoch. Every peak attributes to different neural features, starting from beta in anterior superior temporal gyrus (aSTG) and posterior superior temporal gyrus (pSTG) to HG1, gamma, beta, HG2, etc. frequency bands in all different locations of the brain in chronological order. Even when it is not for the highest accuracy during the whole period, the peak accuracy neural feature in each language processing period could represent the optimal decoding feature for each syntactic, semantic, and unification process of speech perception. Figure 4B elaborates on figure 4A for detailed tracking of the frequency bands that yield the highest accuracy at each time window. The trend line is presented to aid the chronological understanding. The line only connects the bottom value when there are more than one frequency band at a time window, but this does not reflect any priority between plotted frequency bands.

3.3. Cross-subject neural feature comparison

Since individual subjects have different electrode configurations due to clinical requirements, the neural features that yield the highest accuracy are bound to be different subject by subject in spectral, spatial, and temporal aspects. However, to elicit a grain of common pattern across the participants, all the frequency bands that yielded the highest accuracy at a time window were listed for each subject. The three columns do not represent any priority; all listed frequency bands at each time window represent the same classification accuracy. Rows marked in yellow are either 1) absolutely same frequency bands at a given time window across all patients, or 2) relatively same frequency tendencies are sustained for two or more time windows across all patients. An example for the first case would be time window #47 and #51 in table 1. Time windows #13-15 and #33-35 are examples for the second case. The line below time window #38 marks the offset of the word segment '무엇입니까? (which)' for questions 3 and 4. Yellow rows beyond time window #38 could also be reasonable candidates for cross-subject relevant neural features because no further visual or auditory stimuli are given to the subject by time window 60. Therefore, yellow rows after the word segment offset could represent neural features that engage in the late-unification processing of the perceived speech segment.

4. Discussion

The analysis in this study mainly focuses on discerning brain activities generated when processing phonetically controlled speech containing two semantically opposing categories using LDA classifiers. Training and testing features for the LDA decoder model vary in temporal, spatial, and spectral aspects. Decoding accuracies were properly reported using adequate validation methods such as 4-fold cross-validation. Furthermore, based on the neurophysiological account for temporal processing of speech perception, lists of potential neural figures were generated that best represent syntactic, semantic, and unification of classifying two semantic categories of '± life' and '± body part'. Such a suggestion is based on the premise that the neural features that yield the highest classification accuracy embody the most distinctive neural characteristics responsible for separating the semantic categories. Temporal and spectral information of such neural features that bear potential for truly discriminating two semantic categories are provided in table 1. The yellow rows indicate some of the most potent neural feature candidates for semantic speech classification in the given speech task. Due to clinical constraints, those listed neural features might come from electrode channels in different brain areas. This is one of the inherent difficulties when analyzing intracranial EEG data. However, the fact that there are several matching time windows with the specific frequency band that yields high classification accuracy or at least maintain the same frequency band tendency is highly notable, especially for the features that are located within a specific time range that is known to process syntactics, semantics, and unification (100-150ms, 250-500ms, and 500-750ms each). In addition, the study verified that the speech perception is processed in a constant cycle of the temporo-parietal lobe and inferior frontal lobe for constant memory-unification interaction by identifying the brain regions responsible for the highest accuracy in chronological order. Figure 4 clearly shows the feature location alternates between posterior temporal and inferior frontal lobe, occasionally including the hippocampus depth electrodes.

One of the novelty and limitations of this study at the same time is that the speech task aimed to imitate the natural conversational question-and-answer type speech semantic processing, compared to less natural context studies, such as presenting random nonsense phrases. Even though the semantic meanings are on two extremes, decoding characteristic neural features that can discriminate between semantic traits of speech that are both sensible requires meticulous task design. In the future study, the speech task design could be adjusted to ensure a clearer understanding of the desired semantic categories. For example, attentively choosing words used in the task from the level of morphemes or leaving enough silence between words so that the unification and the following phonetic processing do not overlap.

In conclusion, the present study using ECoG signals from the human brain suggests temporal and spectral information of potentially significant neural features that can discern two respective semantic categories of ‘life’ and ‘body part’ perceived during the question-and-answer speech task. Furthermore, tracking the spatial information of such neural features verified the current speech perception theory of temporoparietal and frontal lobe engagement cycle for subserving word meaning comprehension. These findings have implications for further investigation of additional semantic classification features within and outside the current semantic speech task.

References

1. 1Akbari, H., Khalighinejad, B., Herrero, J. L., Mehta, A. D., & Mesgarani, N. (2018). https://doi.org/10.1101/350124
2. Anumanchipalli, G. K., Chartier, J., & Chang, E. F. (2019). Speech synthesis from neural decoding of spoken sentences. *Nature*, *568*(7753), 493–498. https://doi.org/10.1038/s41586-019-1119-1
3. Bhaya-Grossman, I., & Chang, E. F. (2022). Speech Computations of the Human Superior Temporal Gyrus. *Annu Rev Psychol*, *73*, 79–102. https://doi.org/10.1146/annurev-psych-022321-035256
4. Brumberg, J. S., Krusienski, D. J., Chakrabarti, S., Gunduz, A., Brunner, P., Ritaccio, A. L., & Schalk, G. (2016). Spatio-Temporal Progression of Cortical Activity Related to Continuous Overt and Covert Speech Production in a Reading Task. *PLoS One*, *11*(11), e0166872. https://doi.org/10.1371/journal.pone.0166872
5. Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proc Natl Acad Sci U S A*, *110*(7), 2653–2658. https://doi.org/10.1073/pnas.1216827110
6. Chang, E. F., Raygor, K. P., & Berger, M. S. (2015). Contemporary model of language organization: an overview for neurosurgeons. *J Neurosurg*, *122*(2), 250–261. https://doi.org/10.3171/2014.10.JNS132647
7. Chen, L., Goucha, T., Mannel, C., Friederici, A. D., & Zaccarella, E. (2021). Hierarchical syntactic processing is beyond mere associating: Functional magnetic resonance imaging evidence from a novel artificial grammar. *Hum Brain Mapp*, *42*(10), 3253–3268. https://doi.org/10.1002/hbm.25432
8. Chen, Y., Shimotake, A., Matsumoto, R., Kunieda, T., Kikuchi, T., Miyamoto, S., Fukuyama, H., Takahashi, R., Ikeda, A., & Lambon Ralph, M. A. (2016). The 'when' and 'where' of semantic coding in the anterior temporal lobe: Temporal representational similarity analysis of electrocorticogram data. *Cortex*, *79*, 1–13. https://doi.org/10.1016/j.cortex.2016.02.015
9. Fisher, S. E., & Marcus, G. F. (2006). The eloquent ape: genes, brains and the evolution of language. *Nat Rev Genet*, *7*(1), 9–20. https://doi.org/10.1038/nrg1747
10. Flinker, A., Korzeniewska, A., Shestyuk, A. Y., Franaszczuk, P. J., Dronkers, N. F., Knight, R. T., & Crone, N. E. (2015). Redefining the role of Broca's area in speech. *Proc Natl Acad Sci U S A*, *112*(9),

2871–2875. https://doi.org/10.1073/pnas.1414491112
11. Fox, N. P., Leonard, M., Sjerps, M. J., & Chang, E. F. (2020). Transformation of a temporal speech cue to a spatial neural code in human auditory cortex. *Elife*, *9*. https://doi.org/10.7554/eLife.53051
12. Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Hum Brain Mapp*, *31*(3), 448–457. https://doi.org/10.1002/hbm.20878
13. Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063
14. Hagoort, P. (2017). The core and beyond in the language-ready brain. *Neurosci Biobehav Rev*, *81*(Pt B), 194–204. https://doi.org/10.1016/j.neubiorev.2017.01.048
15. Hamilton, L. S., Oganian, Y., & Chang, E. F. (2020). https://doi.org/10.1101/2020.06.08.121624
16. Herff, C., & Schultz, T. (2016). Automatic Speech Recognition from Neural Signals: A Focused Review. *Front Neurosci*, *10*, 429. https://doi.org/10.3389/fnins.2016.00429
17. Hickok, G. (2012). The cortical organization of speech processing: feedback control and predictive coding the context of a dual-stream model. *J Commun Disord*, *45*(6), 393–402. https://doi.org/10.1016/j.jcomdis.2012.06.004
18. Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nat Rev Neurosci*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113
19. Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. https://doi.org/10.1038/nature17637
20. Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, *76*(6), 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014
21. Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Front Psychol*, *2*, 255. https://doi.org/10.3389/fpsyg.2011.00255
22. Kojima, K., Brown, E. C., Rothermel, R., Carlson, A., Fuerst, D., Matsuzaki, N., Shah, A., Atkinson, M., Basha, M., Mittal, S., Sood, S., & Asano, E. (2013). Clinical significance and developmental changes of auditory-language-related gamma activity. *Clin Neurophysiol*,

124(5), 857–869. https://doi.org/10.1016/j.clinph.2012.09.031
23. Kojima, K., Oganian, Y., Cai, C., Findlay, A., Chang, E., & Nagarajan, S. (2021). https://doi.org/10.1101/2020.04.02.022616
24. Li, Y., Tang, C., Lu, J., Wu, J., & Chang, E. F. (2021). Human cortical encoding of pitch in tonal and non-tonal languages. *Nat Commun*, *12*(1), 1161. https://doi.org/10.1038/s41467-021-21430-x
25. Makin, J. G., Moses, D. A., & Chang, E. F. (2020). Machine translation of cortical activity to text with an encoder-decoder framework. *Nat Neurosci*, *23*(4), 575–582. https://doi.org/10.1038/s41593-020-0608-8
26. Martin, S., Brunner, P., Holdgraf, C., Heinze, H. J., Crone, N. E., Rieger, J., Schalk, G., Knight, R. T., & Pasley, B. N. (2014). Decoding spectrotemporal features of overt and covert speech from the human cortex. *Front Neuroeng*, *7*, 14. https://doi.org/10.3389/fneng.2014.00014
27. Martin, S., Iturrate, I., Millan, J. D. R., Knight, R. T., & Pasley, B. N. (2018). Decoding Inner Speech Using Electrocorticography: Progress and Challenges Toward a Speech Prosthesis. *Front Neurosci*, *12*, 422. https://doi.org/10.3389/fnins.2018.00422
28. Martin, S., Millan, J. D. R., Knight, R. T., & Pasley, B. N. (2019). The use of intracranial recordings to decode human language: Challenges and opportunities. *Brain Lang*, *193*, 73–83. https://doi.org/10.1016/j.bandl.2016.06.003
29. Matchin, W., & Hickok, G. https://doi.org/10.31234/osf.io/6394f
30. Matchin, W., & Hickok, G. (2020). The Cortical Organization of Syntax. *Cereb Cortex*, *30*(3), 1481–1498. https://doi.org/10.1093/cercor/bhz180
31. Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. *Eur J Neurosci*, *48*(7), 2609–2621. https://doi.org/10.1111/ejn.13748
32. Milsap, G., Collard, M., Coogan, C., Rabbani, Q., Wang, Y., & Crone, N. E. (2019). Keyword Spotting Using Human Electrocorticographic Recordings. *Front Neurosci*, *13*, 60. https://doi.org/10.3389/fnins.2019.00060
33. Moses, D. A., Leonard, M. K., Makin, J. G., & Chang, E. F. (2019). Real-time decoding of question-and-answer speech dialogue using human cortical activity. *Nat Commun*, *10*(1), 3096. https://doi.org/10.1038/s41467-019-10994-4
34. Moses, D. A., Mesgarani, N., Leonard, M. K., & Chang, E. F. (2016). Neural speech recognition: continuous phoneme decoding using spatiotemporal representations of human cortical activity. *J Neural*

Eng, *13*(5), 056004. https://doi.org/10.1088/1741-2560/13/5/056004
35. Popham, S. F., Huth, A. G., Bilenko, N. Y., Deniz, F., Gao, J. S., Nunez-Elizalde, A. O., & Gallant, J. L. (2021). Visual and linguistic semantic representations are aligned at the border of human visual cortex. *Nat Neurosci*, *24*(11), 1628-1636. https://doi.org/10.1038/s41593-021-00921-6
36. Rabbani, Q., Milsap, G., & Crone, N. E. (2019). The Potential for a Speech Brain-Computer Interface Using Chronic Electrocorticography. *Neurotherapeutics*, *16*(1), 144-165. https://doi.org/10.1007/s13311-018-00692-2
37. Rupp, K., Roos, M., Milsap, G., Caceres, C., Ratto, C., Chevillet, M., Crone, N. E., & Wolmetz, M. (2017). Semantic attributes are encoded in human electrocorticographic signals during visual object recognition. *Neuroimage*, *148*, 318-329. https://doi.org/10.1016/j.neuroimage.2016.12.074
38. Scott, S. K. (2019). From speech and talkers to the social world: The neural processing of human spoken language. *Science*, *366*(6461), 58-62. https://doi.org/10.1126/science.aax0288
39. Sjerps, M. J., Fox, N. P., Johnson, K., & Chang, E. F. (2019). Speaker-normalized sound representations in the human auditory cortex. *Nat Commun*, *10*(1), 2465. https://doi.org/10.1038/s41467-019-10365-z
40. Strauss, A., Kotz, S. A., Scharinger, M., & Obleser, J. (2014). Alpha and theta brain oscillations index dissociable processes in spoken word recognition. *Neuroimage*, *97*, 387-395. https://doi.org/10.1016/j.neuroimage.2014.04.005
41. Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., & Francart, T. (2018). https://doi.org/10.1101/246660
42. Sato,N. (2021)Frequency-dependent Cortical Interactions during Semantic Processing: An Electrocorticogram Cross-spectrum Analysis Using a Semantic Space Model. *Cerebral Cortex, 31, 4329-4339*
43. Suh MW, Lee HJ, Kim JS, Chung CK, Oh SH. Speech experience shapes the speechreading network and subsequent deafness facilitates it. *Brain*. 2009 Oct;132(Pt 10):2761-71. doi: 10.1093/brain/awp159. Epub 2009 Jun 16. PMID:19531532.
44. Kim CH, Kim JH, Chung CK, Kim JS, Lee JM, Lee SK. Localization of Broca's Area Using Functional MR Imaging: Quantitative Evaluation of Paradigms. *J KoreanNeurosurg Soc*. 2009 Apr;45(4):219-23. doi: 10.3340/jkns.2009.45.4.219. Epub 2009 Apr 30. PMID: 19444347; PMCID: PMC2682117.

45. Lee SY, Kim JS, Chung CK, Lee SK, Kim WS. Assessment of language dominance by event−related oscillatory changes in an auditory language task: magnetoencephalography study. *J Clin Neurophysiol*. 2010 Aug;27(4):263−9. doi:10.1097/WNP.0b013e3181eaaf3c. PMID: 20634707.
46. Kim JS, Chung CK. Language lateralization using MEG beta frequency desynchronization during auditory oddball stimulation with one−syllable words. *Neuroimage*. 2008 Oct 1;42(4):1499−507. doi:10.1016/j.neuroimage.2008.06.001. Epub 2008 Jun 13. PMID: 18603004.
47. Kim CH, Chung CK, Koo BB, Lee JM, Kim JS, Lee SK. Changes in language pathways in patients with temporal lobe epilepsy: diffusion tensor imaging analysis of the uncinate and arcuate fasciculi. *World Neurosurg*. 2011 Mar−Apr;75(3−4):509−16. doi:10.1016/j.wneu.2010.11.006. PMID: 21600505.
48. Kim JH, Lee JM, Kang E, Kim JS, Song IC, Chung CK. Functional reorganization associated with semantic language processing in temporal lobe epilepsy patients after anterior temporal lobectomy : a longitudinal functional magnetic resonance image study. *J Korean Neurosurg Soc*. 2010 Jan;47(1):17−25.doi:10.3340/jkns.2010.47.1.17. Epub 2010 Jan 31. PMID: 20157373; PMCID: PMC2817510.

List of Figures

Question	Possible Answers
Q1. Which of the above is alive? (다음 중 살아있는 것은 무엇입니까?)	Dog, human, bear (강아지, 사람, 곰)
Q2. Which of the above is not alive? (다음 중 살아있지 않은 것은 무엇입니까?)	Snowman, hat, book (눈사람, 모자, 책)
Q3. Which of the above is a part of a body? (다음 중 몸의 일부인 것은 무엇입니까?)	Eye, nose, hand (눈, 코, 손)
Q4. Which of the above is not a part of a body? (다음 중 몸의 일부가 아닌 것은 무엇입니까?)	Flower, cup, gun (꽃, 컵, 총)

Figure 1. Question−and−answer speech task design.
The task was designed to investigate brain activity during proactive semantic processing when asked specific questions containing basic semantic meaning (compared to passively processing auditory speech sound and producing meaningless sound). Questions were selected as above so that the questions contain basic, non−abstract semantic meaning (alive/ not alive, is a body part/ is not a body part) that can be understood by the subject regardless of the patients' age or gender

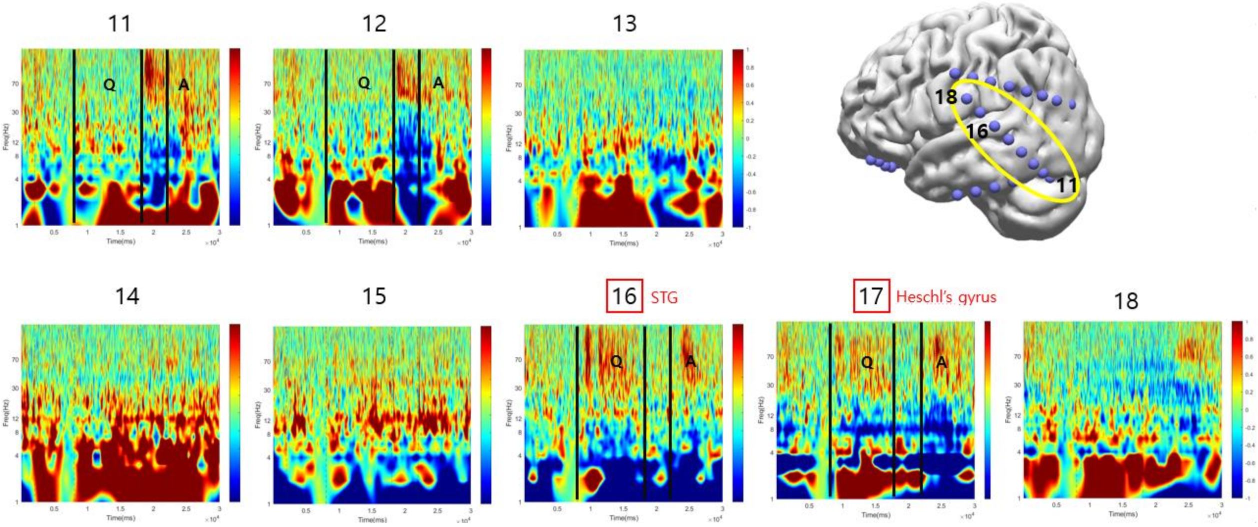


Figure 2. Time-frequency analysis showing spectral dynamics during the speech task by continuous wavelet transform analysis.
Subject 7 had optimal electrode placement for investigating change in spectral dynamics in the temporal lobe and adjacent regions. Speech task brain signals were averaged across trials and plotted for mean time-frequency representation.

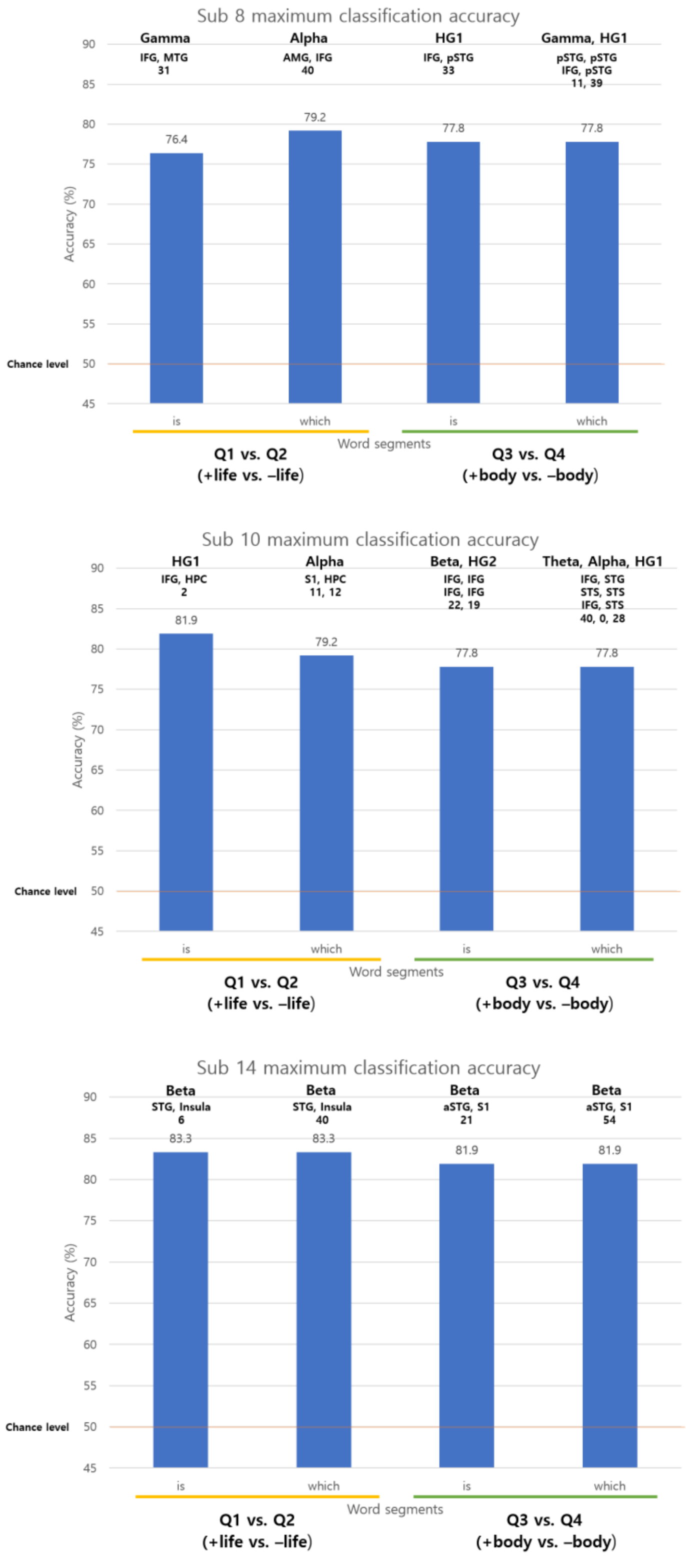


Figure 3. Maximum classification accuracy per each subject
Classification accuracy generated using linear discriminant analysis (LDA)

classifier. Q1 vs. Q2 refers to classifying between the semantic category of +life and −life, by classifying neural signals from each time window (50ms each) of each word segment '것은 (is)', and '무엇입니까 (which)'. The same applies for classifying the second semantic category of +body part and − body part. At the top of each bar graph are spectral, spatial, and temporal information of the neural feature that yielded the highest classification accuracy. Chance level is marked with red horizontal line at 50%.

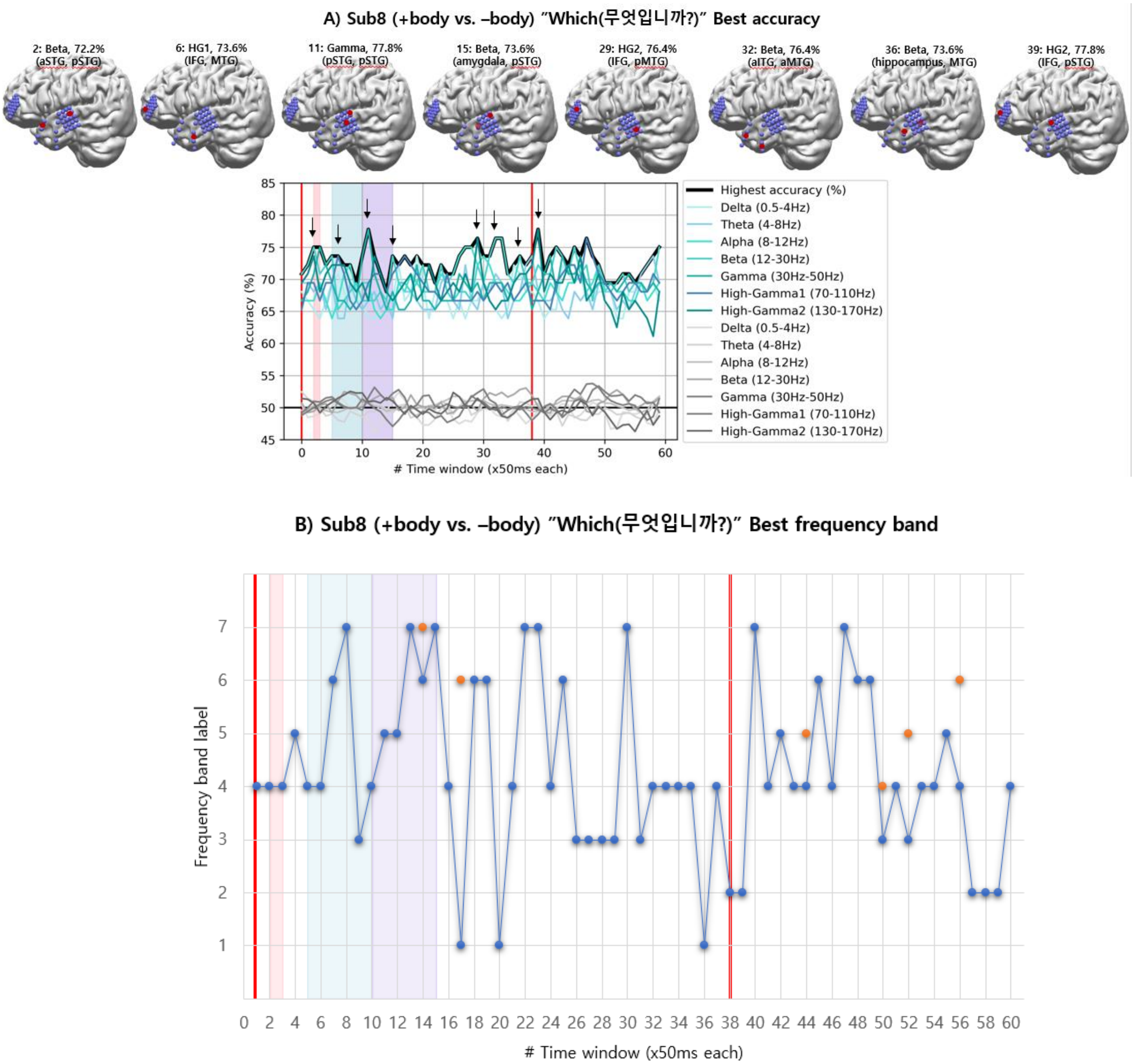


Figure 4. Chronological plot of decoding accuracy for sub8, classifying "무엇입니까?(Which)" word segment from Q3 and Q4. X axis refers to the index number of the time window, and the red vertical lines indicate the onset and offset of the word segment.

A) Superposed plot of best classification accuracy from all frequency ranges (Delta, theta, alpha, beta, gamma, high−gamma1, high−gamma2, all in light blue shades). Below the best accuracies are mean classification accuracies which are about chance level (50%) (all frequency ranges in gray shades). Pink back ground refers to time epoch known for the syntactic processing onset, blue back ground for the semantic processing onset, and the purple background for the unification onset. Spatial neural feature information for selected peaks marked with black arrows are illustrated above the graph. The red dots on the brain figures locates the two neural features that together yielded the best accuracy.

B) Elaboration on frequency band rankings for each time window. 1−7 on the y axis indicates Delta – HG2 frequency bands. More than one frequency bands were marked if there were multiple features that yielded the same value of the highest accuracy. The trace line only connects the bottom marks only to show the temporal concept of the graph. Marks in orange bears the same importance as the marks in blue, they are different in color to show that there are more than one feature in the specific time window.

	Sub8		Sub10			Sub14		
me windo	fq. Band	fq. Band	fq. Band	fq. Band	fq. Band	fq. Band	fq. Band	fq. Band
1	4		3			5		
2	4		3			1	2	7
3	4		6			1	2	5
4	5		6			6		
5	4		6			2	7	
6	4		3			7		
7	6		5			2		
8	7		3			3		
9	3		1			2	3	
10	4		5			2	3	6
11	5		5			7		
12	5		2	3	7	2		
13	7		6			2		
14	6	7	6			2		
15	7		6	7		2		
16	4		7			2	4	
17	1	6	7			2		
18	6		7			5	7	
19	6		7			4		
20	1		2			4		
21	4		7			4		
22	7		4			7		
23	7		3			7		
24	4		3			5		
25	6		3			6		
26	3		3			6		
27	3		6			4		
28	3		3			7		
29	3		6			5		
30	7		1			5		
31	3		1			4		
32	4		1	4		6		
33	4		4			4		
34	4		5			7		
35	4		5			7		
36	1		5			7		
37	4		3	4		7		
38	2		7			7		
39	2		4			7		
40	7		2	4		7		
41	4		2			3		
42	5		2			3		
43	4		5			3		
44	4	5	6			4	6	
45	6		7			4		
46	4		6			4		
47	7		7			4	7	
48	6		6			2		
49	6		4			5		
50	3	4	3			5		
51	4		4			4		
52	3	5	2			1		
53	4		3			4		
54	4		2			4		
55	5		3			4		
56	4	6	6			4		
57	2		3			4		
58	2		2			4		
59	2		5			2		
60	4		3			1	2	4

Table 1. Cross−subject best classification feature comparison.

The numbers from 1 to 7 entered in the columns named fq.Band refers to each frequency bands from delta to HG2. For classifying Q3 vs. Q4 (+body vs. −body) by word segment '무엇입니까? (Which)', the spectral frequency bands that yielded the highest classification accuracy at specific time windows are listed. Rows in yellow color mark either 1) absolutely same frequency bands at a given time window across all patients, or 2) relatively same frequency tendencies are sustained for two or more time windows across all patients.

Supplementary information

	[Q1, Q2: Alive/not alive]				[Q3, Q4: a body part/not a body part]			
trial/session	1	2	5	6	3	4	7	8
1	강아지	사람	눈사람	책	눈	코	꽃	총
2	곰	곰	사람	강아지	손	손	코	눈
3	눈사람	모자	강아지	사람	꽃	컵	눈	코
4	모자	강아지	모자	눈사람	컵	눈	컵	꽃
5	강아지	곰	눈사람	모자	눈	손	꽃	컵
6	책	눈사람	곰	곰	총	꽃	손	손
7	곰	사람	책	강아지	손	코	총	눈
8	모자	책	강아지	곰	컵	총	눈	손
9	사람	모자	사람	책	코	컵	코	총
10	눈사람	강아지	모자	사람	꽃	눈	컵	코
11	곰	눈사람	곰	곰	손	꽃	손	손
12	책	책	책	모자	총	총	총	컵
13	강아지	강아지	곰	책	눈	눈	손	총
14	눈사람	책	모자	눈사람	꽃	총	컵	꽃
15	사람	눈사람	사람	강아지	코	꽃	코	눈
16	책	사람	강아지	모자	총	코	눈	컵
17	모자	곰	눈사람	사람	컵	손	꽃	코
18	사람	모자	책	눈사람	코	컵	총	꽃

Table S1. Detailed question−and−answer speech task session/trial information. There are 12 possible answers: 강아지, 사람, 곰, 눈사람, 모자, 책, 눈, 코, 손, 꽃, 컵, 총) After finishing all sessions 1−8, each answer has appeared 12 times. During sessions 1, 2, 5, 6: question 1 and question 2 appears 36 times each, and during sessions 3, 4, 7, 8: question 3 and question 5 appears 36 times each.

Subject number	Sex	Age	Conducted semantic sessions	Epilepsy type	Hemisphere	Number of electrodes
1	F	36	Sem1−8	TLE	Left	46
2	M	21	Sem1−8	Generalized tonic−clonic	Right	24
3	F	29	Sem1−8	TLE−	Right	54
4	M	49	Sem1−8	−	Left	32
5	M	36	Sem1−5, 7	TLE	Left	23
6	M	44	Sem2−8	TLE	Right	42
7	F	31	Sem1−8	TLE	Left	36
8	M	27	Sem1−8	TLE	Left	58
9	M	20	Sem1−8	FLE	Left	68
10	M	48	Sem1−8	TLE	Left	68
11	F	20	Sem1−8	TLE	Left	72
12	M	39	Sem1−8	TLE	Left	82
13	F	25	Sem1−8	TLE	Left	60
14	M	50	Sem1−8	TLE	Left	76

Table S2. Patient information

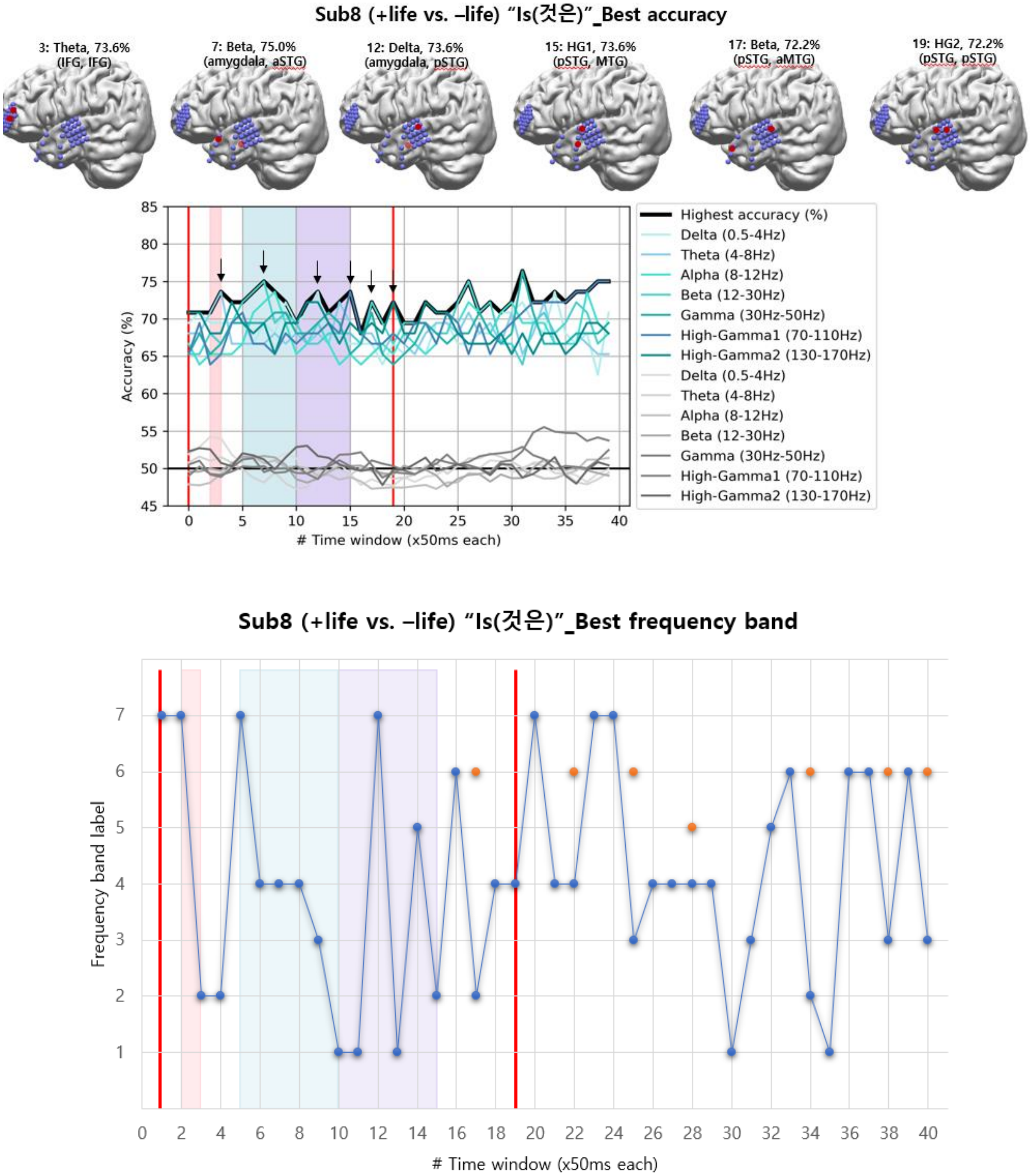
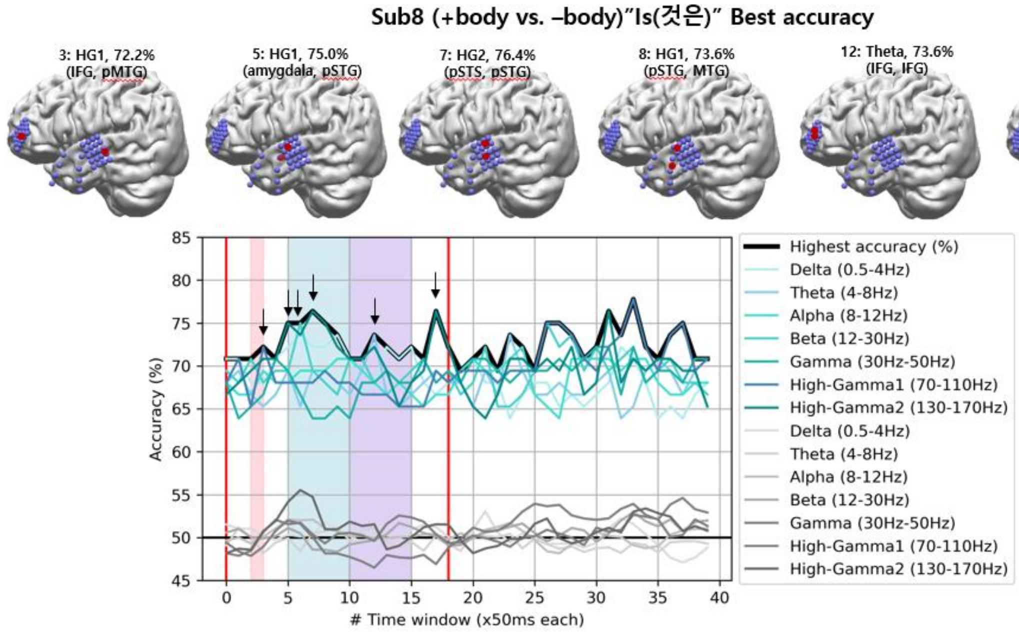


Figure S1. Chronological plot of decoding accuracy and best frequency bands for sub8, classifying "것은 (is)" for classifying Q1 and Q2.



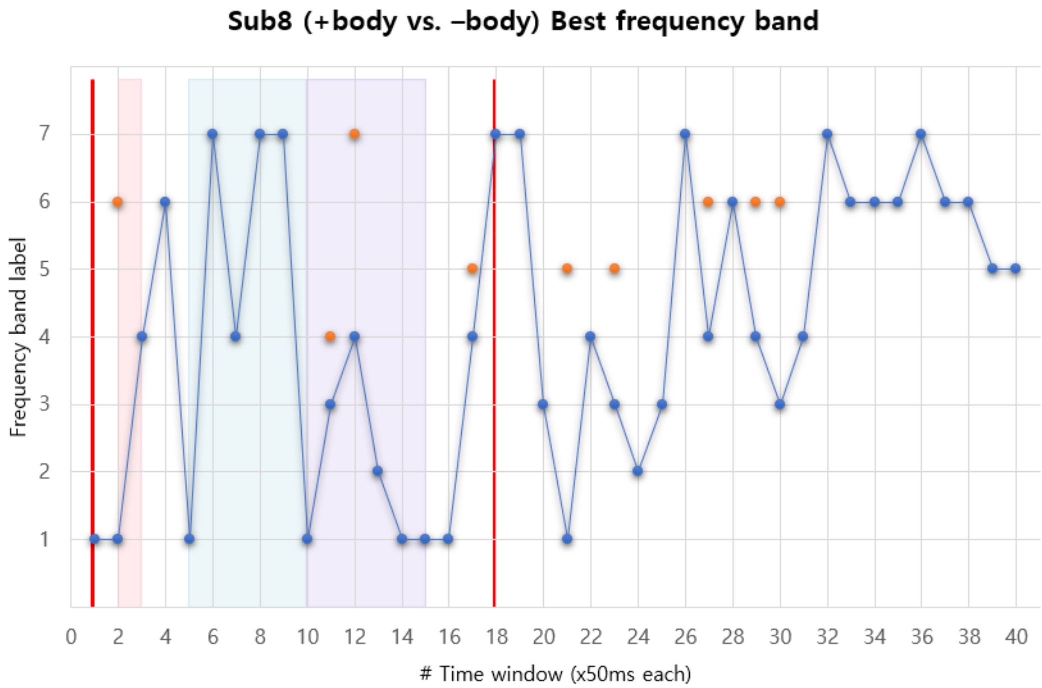


Figure S2. Chronological plot of decoding accuracy and best frequency bands for sub8, classifying "것은 (is)" for classifying Q3 and Q4.

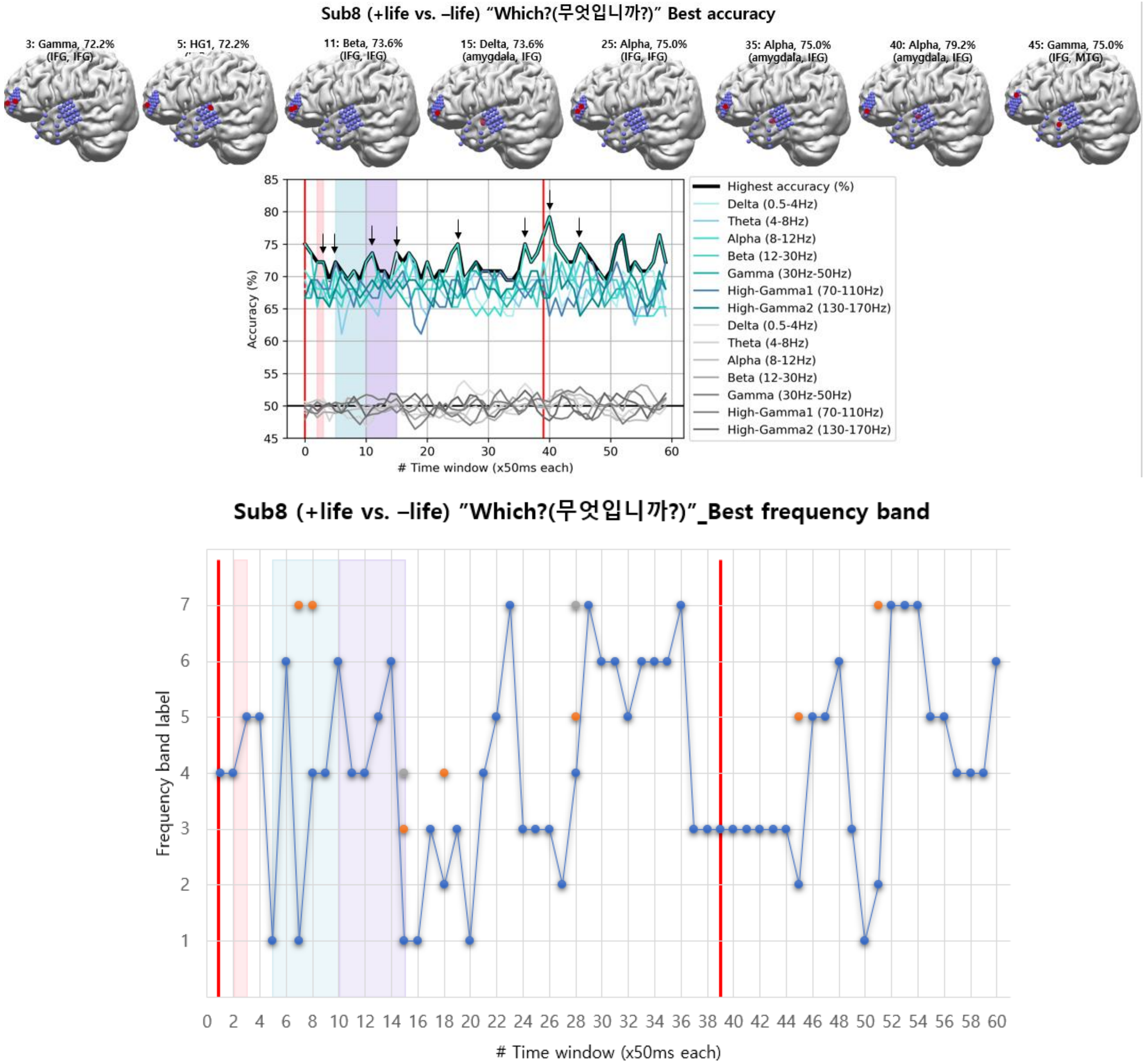


Figure S3. Chronological plot of decoding accuracy and best frequency bands for sub8, classifying "무엇입니까? (which)" for classifying Q1 and Q2

	Sub8		Sub10			Sub14		
me windo	fq. Band	fq. Band	fq. Band	fq. Band	fq. Band	fq. Band	fq. Band	fq. Band
1	7		6			5		
2	7		6			5		
3	2		6			4		
4	2		6			4		
5	7		4	6		6		
6	4		6			4		
7	4		6			4		
8	4		4	6		4		
9	3		1			4		
10	1		4			4		
11	1		4			4		
12	7		4	6		6		
13	1		1			4		
14	5		4			7		
15	2		4			7		
16	6		7			6	7	
17	2	6	4			6		
18	4		4			6	7	
19	4		4			4		
20	7		5			4		
21	4		4	5	6	4		
22	4	6	4			4	7	
23	7		4			4	7	
24	7		4			6		
25	3	6	3	5		7		
26	4		4			7		
27	4		4			7		
28	4	5	4			7		
29	4		3	4	5	7		
30	1		5			4		
31	3		6			4		
32	5		6			4		
33	6		6			4		
34	2	6	2			4		
35	1		4			7		
36	6		4			7		
37	6		4			4		
38	3	6	6			4		
39	6		6			4		
40	3	6	6			4		

Table S3. Cross−subject best classification feature comparison for classifying Q1 vs. Q2 (+life vs. −life) by word segment '것은 (is)'

	Sub8		Sub10			Sub14	
me windo	fq. Band	fq. Band	fq. Band	fq. Band	fq. Band	fq. Band	fq. Band
1	1		3	7		6	
2	1	6	4	5		6	
3	4		4			7	
4	6		3			6	
5	1		3	4		5	
6	7		6			6	
7	4		5			4	
8	7		3	4		4	
9	7		2			4	
10	1		1			4	
11	3	4	2			6	
12	4	7	1			3	
13	2		7			2	3
14	1		4			4	
15	1		6			6	
16	1		4			6	
17	4	5	3	4		6	
18	7		4			4	
19	7		7			4	
20	3		7			4	
21	1	5	6			4	
22	4		4			4	
23	3	5	4			5	
24	2		4			6	
25	3		4			6	
26	7		6			3	
27	4	6	7			4	7
28	6		2			3	
29	4	6	5			5	
30	3	6	3	5		5	
31	4		6			5	
32	7		1			2	5
33	6		2			5	
34	6		1	7		5	
35	6		2			5	
36	7		3			5	
37	6		3			5	
38	6		2			3	
39	5		2	3	4	5	
40	5		2			5	

Table S4. Cross−subject best classification feature comparison for classifying Q3 vs. Q4 (+body vs. −body) by word segment ‘것은 (is)’

me windo	Sub8 fq. Band	Sub8 fq. Band	Sub8 fq. Band	Sub10 fq. Band	Sub10 fq. Band	Sub14 fq. Band	Sub14 fq. Band	Sub14 fq. Band
1	4			3		2		
2	4			6		2	5	
3	5			4		7		
4	5			5		2		
5	1			4		2	6	
6	6			3		2	3	7
7	1	7		3		7		
8	4	7		3		2	3	
9	4			3		3		
10	6			4		3		
11	4			3		3	4	
12	4			3		2	5	
13	5			3		1		
14	6			2		6		
15	1	3	4	5		2	6	
16	1			1	2	6	7	
17	3			2		2		
18	2	4		5	6	2		
19	3			5		1		
20	1			5		7		
21	4			1	7	6		
22	5			5		3		
23	7			7		2	3	5
24	3			7		2		
25	3			3	4	3		
26	3			3		4	5	
27	2			5	7	5		
28	4	5	7	5	6	5		
29	7			6		2	6	
30	6			3		6		
31	6			7		2		
32	5			1	7	7		
33	6			4		7		
34	6			4		7		
35	6			3	4	7		
36	7			7	4	7		
37	3			4		4	7	
38	3			1	7	7		
39	3			6		7		
40	3			5	7	7		
41	3			5		4		
42	3			5		4		
43	3			5		4		
44	3			5		4		
45	2	5		5		4		
46	5			2		4		
47	5			2		7		
48	6			2		7		
49	3			4		7		
50	1			4		7		
51	2	7		7		7		
52	7			5	7	7		
53	7			7		7		
54	7			6		4		
55	5			6	7	4		
56	5			6		4		
57	4			6	7	4		
58	4			6		4		
59	4			7		5		
60	6			1		4		

Table S5. Cross−subject best classification feature comparison for classifying Q1 vs. Q2 (+life vs. −life) by word segment '무엇입니까? (which)'

국문 초록

음성 의미 지각시의 고등 언어 성분 처리 디코딩

인간의 고등 성분 언어 처리와 관련한 두뇌 활동을 해독하는 연구는 신경언어학 분야에서도 아직 깊이 연구되지 않은 분야 중 하나이다. 침습적 전극을 통해 얻은 뇌피질 뇌파를 이용한 대부분의 언어 디코딩 연구는 음소나 음절 수준의 하위 언어 성분에서 진행되어 왔고, 통사나 의미와 같은 고등 언어 성분에 대한 디코딩 연구는 드물다. 드물게 진행된 고등 언어 성분 디코딩 연구 또한 대다수가 시각적 언어 처리를 연구한 결과들이며, 소리 언어 디코딩 연구는 태동 단계에 머무르고 있다.

본 연구는 소리 언어 지각시의 두뇌 활성을 분석하여 그 처리 과정의 뇌파 신호 특성을 규명하고자 한다. 특히 인간 음성 언어의 하위 구성 성분보다는 통사와 의미 위주의 고등 구성 성분을 처리하는 데에 관여하는 뇌파의 시간적, 주파수적, 공간적 특성에 집중하여 분석을 진행하였다. 언어 처리의 주된 세 가지 요소는 ‘음소 (phonetics)’, ‘통사 (syntactics)’, ‘의미 (semantics)’라는 점을 고려하여, 음소 소준의 뇌파 활동을 통제할 수 있는 실험 패러다임을 구상하였으며, 구체적으로는 두개의 다른 의미 범주 (생명, 신체)에 대해서 묻는 음소적으로 동등한 단어가 포함된 질문을 들려준 후 의미를 파악해 대답하는 과정의 뇌파를 기록하는 실험을 진행하였다.

뇌파 신호는 경막하 전극 삽입술 (Electrocorticography, ECoG)을 통해 14명의 뇌전증 환자로부터 침습적 방식으로 측정되었다. 뇌파 디코딩 분석에는 피험자의 두뇌가 옳은 방식으로 처리한 고등 언어 성분이 반영된 실험만을 포함하기 위해서 모든 실험에서 옳은 대답을 한

세 명의 환자만을 대상으로 하여 분석을 진행하였다. 디코딩 분석 결과 세 명의 환자에 걸쳐 핵심 단어 (‘것은’ , ‘무엇입니까’) 음성 지각 이후 특정 시간대에서 특정 주파수대의 뇌파가 양 극단의 의미를 높은 수준의 정확도(%)로 분류하는 데에 사용될 수 있다는 것을 밝혔다. 또한 이러한 높은 정확도를 기록한 뇌파의 특성에는 모든 환자에 걸쳐 절대적 혹은 상대적 트렌드가 관찰되며, 관찰되는 뇌파의 공간적 특성은 현재 통용되는 신경언어학적 언어 처리 모델이 설명하는 음성 언어 처리 방식과 일맥상통함을 밝혔다.

키워드: 인간 뇌, 의미론, 음성 지각, 경막하 전극 삽입술, 디코딩, 분류

학번: 2020-22887