



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

A Doctor of Philosophy Dissertation

연속 음성 유발 전위 기반  
언어인지도 예측 모델 개발

Prediction of speech intelligibility using speech-  
evoked cortical response

The Graduate School  
of the University of Ulsan

Department of  
Biomedical Engineering

Youngmin Na

Prediction of speech intelligibility using speech-  
evoked cortical response

Supervisor: Jihwan Woo

A Dissertation

Submitted to

the Graduate School of the University of Ulsan

In partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

by

Youngmin Na

Department of Biomedical Engineering

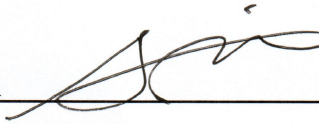
University of Ulsan, Korea

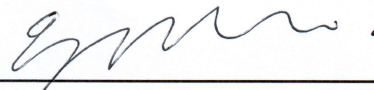
February 2022

Prediction of speech intelligibility using speech-evoked  
cortical response

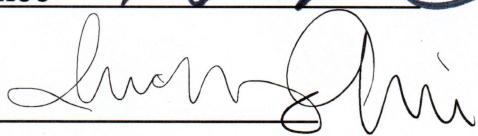
This certifies that the dissertation of  
Youngmin Na is approved.

Committee Chair Dr. Kyo-in Koo 

Committee Member Dr. Sungmin Kim 

Committee Member Dr. Jihwan Woo 

Committee Member Dr. Young-joon Chee 

Committee Member Dr. Inyong Choi 

Department of Biomedical Engineering

University of Ulsan, Korea

February 2022



# Contents

[Abstract].....	VI
[List of Figures].....	VIII
[List of Tables] .....	X
<b>Chapter 1. Introduction</b> .....	1
1.1 Auditory processing .....	2
1.2 Auditory ability assessment.....	3
1.3 Auditory evoked potential (AEP) .....	4
1.4 Speech intelligibility .....	5
1.5 Neural tracking.....	6
1.6 Phonemic information .....	7
1.7 Research goal.....	9
<b>Chapter II. Phonetic-level feature extraction and evaluation for speech intelligibility prediction</b> .....	10
2.1 Methods .....	12
2.1.1 Participants.....	12
2.1.2 Behavioral test.....	12
2.1.3 Stimuli and procedure .....	13
2.1.4 EEG recording and processing.....	14
2.1.5 Continuous speech-evoked potentials .....	14
2.2 Results.....	16
2.2.1 CSEP in sensor level .....	16

2.2.2	CSEP at source level .....	19
2.3	Conclusion & Discussion.....	21
<b>Chapter III. Predicting speech intelligibility using deep learning model.....</b>		<b>23</b>
3.1	Methods .....	25
3.1.1	Behavioral test.....	25
3.1.2	EEG recording and procedure .....	25
3.1.3	Continuous speech-evoked potential with speech feature.....	25
3.1.4	Data augmentation .....	26
3.1.5	Architecture of speech intelligibility prediction model.....	27
3.2	Results.....	28
3.2.1	Speech intelligibility prediction model performance .....	28
3.2.2	Occlusion sensitivity map .....	30
3.3	Conclusion & Discussion.....	31
<b>Chapter VI. General conclusion.....</b>		<b>32</b>
<b>Reference.....</b>		<b>34</b>

[Abstract]

# Prediction of speech intelligibility using speech-evoked cortical response

Youngmin Na

Department of Biomedical Engineering  
The Graduate School of University of Ulsan

This study aimed to develop the deep learning model to assess speech intelligibility (SI) objectively without attention from continuous speech-evoked potential (CSEP). The CSEP extracted in this study was the temporal response function of neural tracking. The neural tracking here is a mathematical approach to quantify how well the entrained activity is aligned to speech features. While a popular speech feature for neural tracking is the speech envelope, phoneme information is crucial to understand the speech. In Chapter II, the phoneme onset time as an event cue was used for phoneme-based neural tracking. The phoneme CSEP was validated using the natural and 4-channel vocoded conditions. The CSEP using phoneme onset neural tracking revealed SI differences at the N1-P2 complex. In other words, phoneme onset time can represent the degree of speech intelligibility.

SI prediction model was developed with speech features of temporal envelope and phoneme information. The SI prediction deep learning model was trained using the features of ERPs, envelope-based CSEPs (ENV), phoneme-based CSEPs (PH), or phoneme-envelope-based CSEPs (PHENV) with the output of behavioral speech intelligibility scores. Data augmentation algorithm was employed to encourage the number of the training dataset. The validation loss of all models decreased during the first two training epochs and saturated thereafter. The deep learning models

were no over-fitted problem. The performances of models were 97.34 (ERP), 99.05 (ENV), 99.87 (PH), and 99.97 % (PHENV), which are comparable to the random chance level of 2.63 %. The results demonstrated that the SI prediction with CSEP could precisely assess speech intelligibility. In addition, the informative electrodes were estimated by using Occlusion sensitivity map. The informative electrodes were language dominant area in the PH and PHENV model.

[Key Words] Speech intelligibility, continuous-speech evoked potential, neural tracking, deep learning model

[List of Figures]

**Figure 1.** A generalized model for bottom-up processing of auditory input (adapted from Edwards, 2007) ..... 2

**Figure 2.** Sentence comprehension processes (Huang, 2015) ..... 3

**Figure 3.** Auditory language comprehension model (Friederici, 2011)..... 4

**Figure 4.** Example of N1-P2 complex based on global field power..... 5

**Figure 5.** Example of ERP to continuous-speech stimulus..... 6

**Figure 6.** Example of neural tracking approach (Weissbart et al., 2020) ..... 7

**Figure 7.** Auditory language comprehension areas each different processing level (Friederici, 2011)..... 8

**Figure 8.** Behavioral speech intelligibility scores across stimulus type ..... 13

**Figure 9.** An example of CSEP extracting process between speech stimulus and EEG signal. **a.** Continuous speech wave form, **b.** phoneme onset impulse train, **c.** EEG signal evoked by continuous speech, **d.** phoneme onset neural tracking ..... 15

**Figure 10.** Grand averaged CSEP of natural (red) and vocoded condition (blue) at left-frontal, left-temporal, and central region..... 17

**Figure 11.** Grand averaged topographies at P1-N1-P2 complex latency..... 18

**Figure 12.** Auditory language processing areas at source level..... 19

**Figure 13.** Grand averaged CSEP in source level and dominance of voxels at N1-P2 complex latency ..... 20

<b>Figure 14.</b> Flow chart of speech intelligibility prediction model from speech signal .....	24
<b>Figure 15.</b> An example of speech features from speech.....	26
<b>Figure 16.</b> Example of data augmentation.....	26
<b>Figure 17.</b> Summary of behavioral SI results. The bar color denoted by stimulus type.....	28
<b>Figure 18.</b> loglized validation loss and accuracy across speech intelligibility models. ....	29
<b>Figure 19.</b> The performance of each speech intelligibility prediction model.....	29
<b>Figure 20.</b> Occlusion sensitivity topographies, color denoted by importance for speech intelligibility prediction.....	30

[List of Tables]

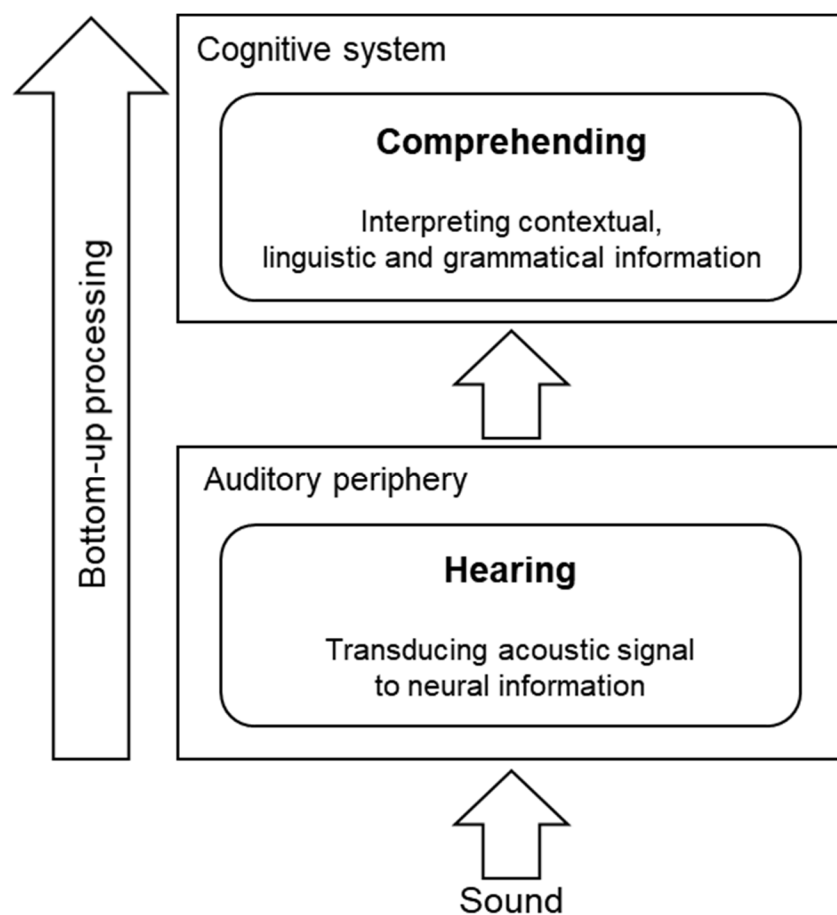
<b>Table I.</b> Selected Korean sentences information from the KS-SL-A.....	14
<b>Table II.</b> Dominance to conditions each language area at N1 <sub>CSEP</sub> .....	20
<b>Table III.</b> Dominance to conditions each language area at P2 <sub>CSEP</sub> .....	21
<b>Table IV.</b> Architecture of speech intelligibility prediction model.....	27
<b>Table V.</b> Informative electrodes for speech intelligibility prediction based on occlusion sensitivity maps .....	30



# **Chapter 1. Introduction**

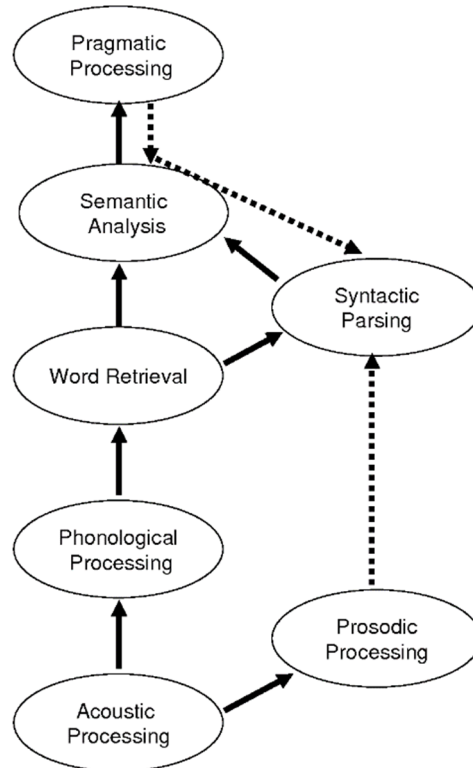
## 1.1 Auditory processing

Auditory processing is a bottom-up phenomenon (Edwards, 2007; Stenfelt and Rönnerberg, 2009). The bottom-up processing begins with sensory data and goes up to the brain's integration of this sensory information (**Figure 1**). The acoustic input is transformed into neural spike trains at cochlear to comprehend speech. The neural spike trains send to the brainstem to sort information before going to the brain. Then this information goes up from the brainstem to the cortex.



**Figure 1.** A generalized model for bottom-up processing of auditory input (adapted from Edwards, 2007)

The brain identifies each word and integrates these words into a structured syntactic and semantic representation (**Figure 2**). Finally, the listener perceives the speech using the representation (Huang, 2015).



**Figure 2.** Sentence comprehension processes (Huang, 2015)

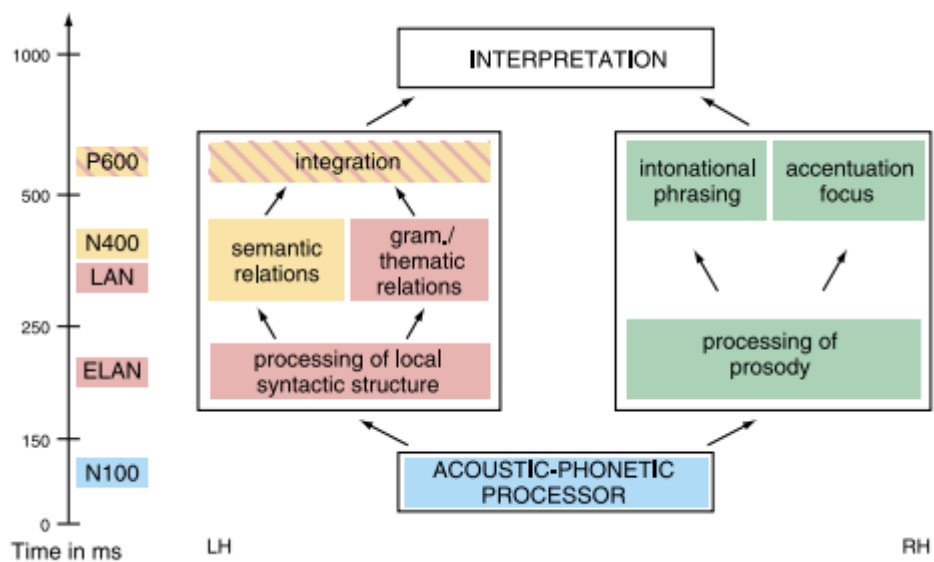
## 1.2 Auditory ability assessment

Hearing-impaired can occur if one of the auditory processes is impaired. Therefore, objectively measuring the functioning of the auditory is important in diagnosis. The gold standard test consists of behavioral and objective measures based on auditory brainstem response (ABR) in the peripheral auditory processing. For example, pure tone audiometry (PTA) is a behavioral test to measure the hearing threshold for each specific frequency. The results of the PTA provide information on the cochlear status. However, for populations that cannot participate in the behavioral test (e.g., infants), an objective measure is ABR using a simple type of sound such as clicks, frequency, or amplitude-modulated tones. Using the ABR, the hearing screening is

performed without the subject’s behavioral response (Chen et al., 1996; Hyde et al., 1990). In the central processing for sentence comprehension, a behavioral test is the gold standard approach in the clinic. Even if researchers have been investigating objective measures based on electroencephalogram (EEG), they are not well-correlated with behavioral tests (Accou et al., 2021; Iotzov and Parra, 2019; Vanthornhout et al., 2018). Therefore, precise objective measurement is required to evaluate the subsystems in the central processing stage.

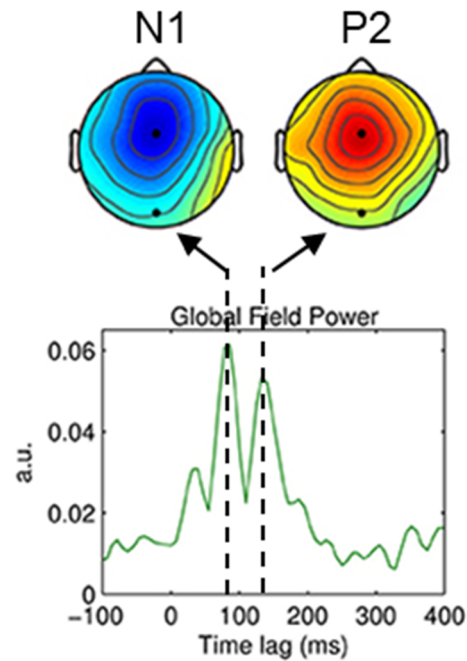
### 1.3 Auditory evoked potential (AEP)

Auditory evoked potential is an event-related potential (ERP) with acoustic stimulus. In the AEP, some components stand for auditory processing levels (**Figure 3**). The N1 of the AEP, a negativity peak around 100 ms after stimulus onset, represents the phoneme identification (Näätänen et al., 1997). The N1 is related to language and reflects the discrimination of auditory categories. Thus, the N1 can be employed to investigate the vowel category perception (Friederici, 2011).



**Figure 3.** Auditory language comprehension model (Friederici, 2011)

In addition, N1 and P2, positivity around 200 ms after stimulus onset, the combination represent auditory processing components (**Figure 4**). The N1-P2 response has been used as an objective predictor of the hearing threshold (Lightfoot, 2016).



**Figure 4.** Example of N1-P2 complex based on global field power

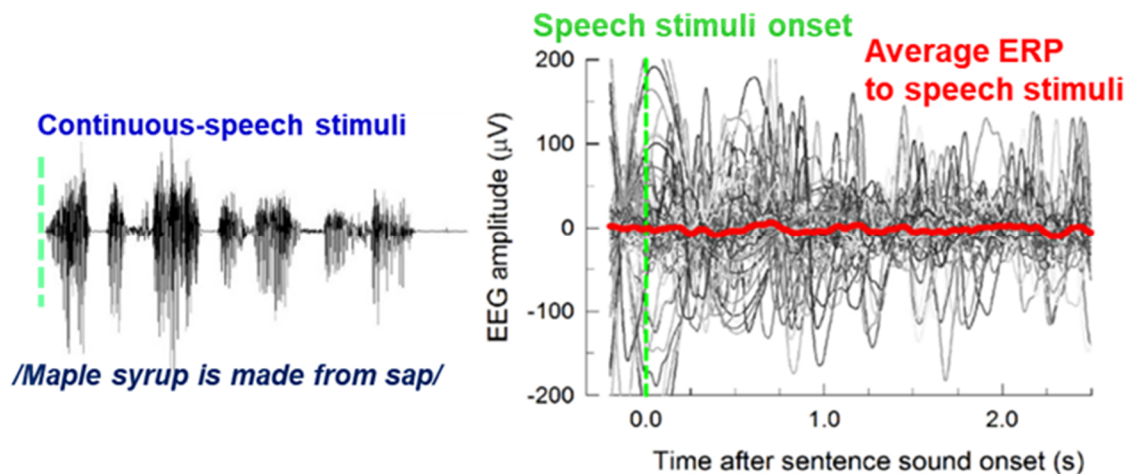
#### 1.4 Speech intelligibility

Speech intelligibility (SI) is an index of the comprehensive level of speech (Venetjoki et al., 2006). The bottom-up whole processing ability can be estimated using the SI. Therefore, SI has also been used to evaluate how well the user fits the auditory prostheses (Kim et al., 2009). Auditory prosthesis provides an excellent opportunity for hearing-impaired patients to rehabilitate the auditory modality. The auditory prosthesis outcome depends on the signal processing strategy and the individual status. A behavioral speech intelligibility test is typically conducted by rating scales how well a listener can comprehend the sentences to evaluate the benefit of auditory prostheses. In particular, a listener is asked to repeat or write down what a listener hears in a

recognition test; then, the speech intelligibility is estimated by scoring the correct number of words. Behavioral-based tests, including rating scales of intelligibility and speech recognition tests, have been widely used in clinics (Ag et al., 2017; D. et al., 1989; Enderby, 1980; Goetz et al., 2008; Healy et al., 2015; Lee, 2016; Robertson, 1982) due to the ease and speed of these approaches. However, the behavioral tests are limited by solely depending on the subject's feelings and required motivation.

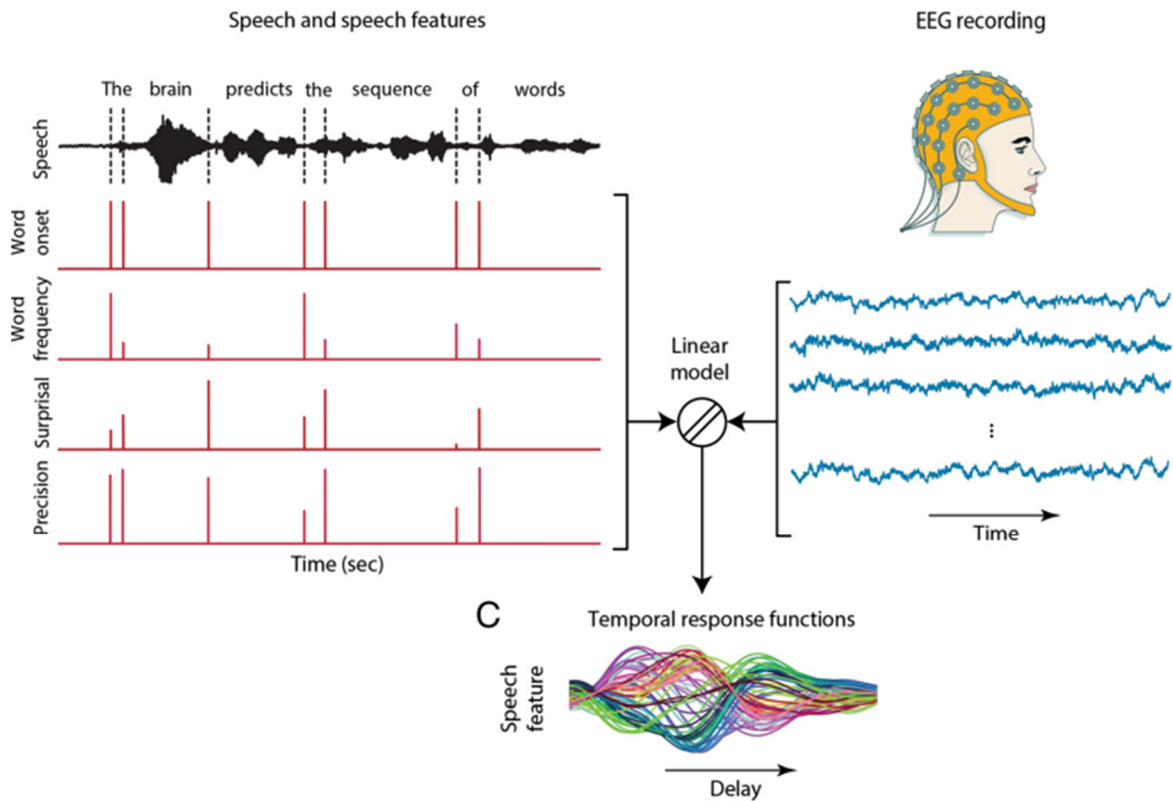
### 1.5 Neural tracking

Although changes in AEP to word or tone stimuli have been recently used to evaluate auditory function objectively, it is vital to objectively diagnose speech intelligibility based on not a single word but the sentences. However, continuous speech-evoked EEG has limited response analysis using the ERP method since it combines ERP by words in a continuous speech, as shown in **Figure 5** (Sanders and Neville, 2003).



**Figure 5.** Example of ERP to continuous-speech stimulus

Recent studies have shown that electroencephalography (EEG) signals to continuous speech are entrained to speech features and reveal a difference corresponding to speech intelligibility (**Figure 6**) (Das et al., 2016; Ding and Simon, 2012; Kong et al., 2015; O’Sullivan et al., 2015). The temporal envelope among the speech features, demonstrated EEG signals were entrained the temporal envelope (Ding and Simon, 2014), has been widely used to investigate the speech intelligibility in the neural tracking literature (Ahissar et al., 2001; Aiken and Picton, 2008; Crosse et al., 2016; Di Liberto et al., 2018; Nourski et al., 2009; O’Sullivan et al., 2015). The components of the neural tracking are analogous to the AEP, reflecting a sequence of neural processing stages within the hierarchy of the auditory system (Davis and Johnsrude, 2003; Di Liberto et al., 2015; Picton, 2013).



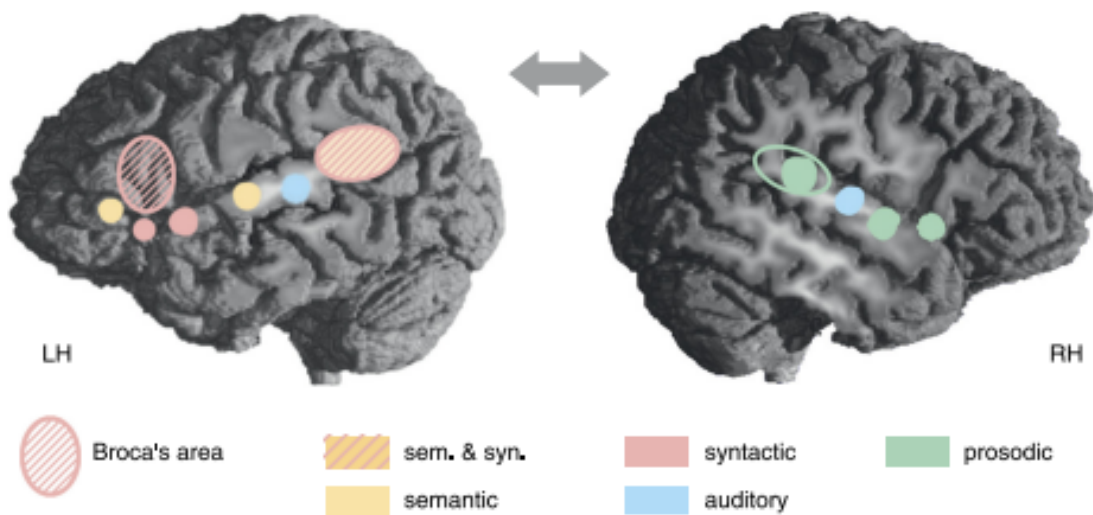
**Figure 6.** Example of neural tracking approach (Weissbart et al., 2020)

### 1.6 Phonemic information

Phonemes can be described as acoustic patterns, but they are also information carriers (Hessler et al., 2013). Speech perception can begin from the phoneme, the information-carrying units, and



then the words can be identified (Shannon, 1948). Investigating continuous speech evoked response based on phonemes could provide an index of lexical processing of speech. (Brodbeck et al., 2018; Donhauser and Baillet, 2020). In addition, Brodbeck et al. reported a response reflecting the incremental integration of phonetic information for word identification, dominantly localized to the left temporal lobe shown in **Figure 7** (Friederici, 2011).



**Figure 7.** Auditory language comprehension areas each different processing level (Friederici, 2011).

## 1.7 Research goal

This dissertation aims to develop an EEG-based speech intelligibility model for the first time to the best of our knowledge. Specifically, the model adopts a passive listening condition to enable the prediction model even in patients who exhibit difficulty with behavioral expression and attention for a long time. The continuous speech evoked potential was extracted from EEG using the phoneme onset information to develop a highly-accurate model. The phoneme-based continuous speech evoked potential was validated by within-subject comparison with a natural and vocoded speech in Chapter II. The speech intelligibility prediction models were developed across speech features using deep learning model and investigated potential various speech features using model performance Chapter III.

## **Chapter II. Phonetic-level feature extraction and evaluation for speech intelligibility prediction**

Recent studies investigating the relationship between SI and neural tracking of the temporal envelope of continuous speech have shown consistent but unexpected results. That is, paradoxically, listeners with poor SI exhibit stronger cortical responses to the temporal envelopes of natural continuous speech (Decruy et al., 2021; Karunathilake et al., 2021). Those studies showed that older adults tracked speech envelopes better than younger ones, even with poorer signal-to-ratio. Interestingly, they showed an increase in cortical amplitude with age. These findings are counterintuitive, as they are inversely related to the effect of profound and long-lasting hearing deficits on auditory sensitivity. A possible explanation for the above findings is that listeners with poor SI employ compensatory mechanisms to account for their poorer speech comprehension (perhaps through exaggerated neural tracking of sensory inputs).

An alternative interpretation for the stronger responses to temporal envelopes associated with poor SI is that tracking temporal envelopes is a bad strategy for speech comprehension that should involve decoding phonetic and linguistic events from speech waveforms. Amplitude increases in speech waveforms do not convey crucial phonetic or linguistic events. Thus, tracking the speech envelope is not necessarily a good measure of speech perception.

Recently, phoneme onsets have been employed to reveal neural responses to continuous speech (Khalighinejad et al., 2017; Liebenthal et al., 2005; Scott et al., 2000), assuming that phoneme onsets form more crucial linguistic events than overall amplitude changes. Using scalp EEG, within-subject comparisons between cortical neural tracking of highly intelligible natural sentences and barely intelligible vocoded sentences were performed by extracting phoneme onset-related responses. We hypothesize that cortical responses to phoneme onsets will be stronger in the natural speech condition than in the vocoded condition.

Although hearing-impaired patients have struggled to attend sound for a long time, previous speech intelligibility prediction research has employed an active listening task. At the high signal-to-noise ratios, tracking the speech is similar between active listening and passive listening which the subjects ignore the stimulus and watch a silent movie instead (Brodbeck and Simon, 2020).

Therefore, passive listening could predict speech intelligibility and fit for hearing-impaired patients.

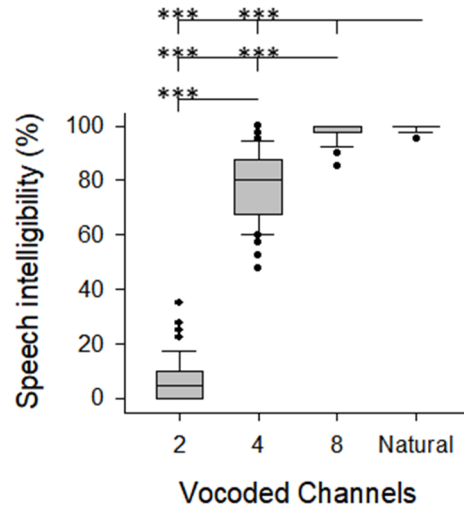
## 2.1 Methods

### 2.1.1 Participants

Fifty normal hearing subjects (25 men and 25 women) participated in this study. All subjects were born at full-term, with no reported health problems. All study procedures were reviewed and approved by the Institutional Review Board of the University of Iowa. All participants signed an informed consent. This study was carried out in accordance with approved guidelines.

### 2.1.2 Behavioral test

The Korean sentence recognition test with natural, 2-, 4-, and 8-channel noise-vocoded sentences were randomly conducted to obtain the behavioral speech intelligibility scores prior to EEG data acquisition. The behavioral speech intelligibility score was evaluated in the Korean sentence recognition test using 41 discrete scores, defined at every 2.5 % between 0 and 100 %. The behavioral speech intelligibility scores of natural, 2-, 4-, and 8-channel noise-vocoded sentences cover the 32 ranges among 41 levels. The SI scores of 2-, 4-channel noise-vocoded conditions are significantly different from the natural condition ( $p < 0.001$ , **Figure 8**). The 4-channel noise-vocoded and natural conditions were employed to compare CSEP.



**Figure 8.** Behavioral speech intelligibility scores across stimulus type

### 2.1.3 Stimuli and procedure

**Table I** shows that ten continuous Korean sentences were selected from the Korean standard sentence lists for adults (KS-SL-A)(Jang et al., 2008). These sentences were normalized and degraded by a 2-, 4-, and 8-channel vocoder to provide a lower SI (Wilson et al., 1991). Sentence duration was < 2.2 s, and the mean of the number of phonemes was 18.6 (std: 3.9). Participants were asked to sit 1 m away from two loudspeakers and watch a silent video while the sentences were played randomly through the loudspeakers. Passive listening tasks with degraded speech (vocoded condition) and clean speech (natural condition) were performed in a soundproof room. Each sentence was randomly repeated 100 times, and the inter-stimulus interval was set to 3 s.

**Table I.** Selected Korean sentences information from the KS-SL-A

	Sentences	Number of phonemes	Duration (s)
1	우체국은 병원 앞에 있어요.	20	2.15
2	당근은 무슨 색입니까?	22	1.84
3	좋아하는 음식이 뭐니까?	21	1.92
4	저녁에 무엇을 먹을까?	19	1.76
5	신발을 벗어 주세요.	17	1.66
6	주차장은 지하에 있습니다.	22	1.92
7	우표 한 장은 얼마입니까?	21	2.08
8	택배가 언제 옵니까?	17	1.65
9	영화는 언제 시작합니까?	21	2.14
10	팔 층을 눌러 주세요.	18	1.62

#### 2.1.4 EEG recording and processing

EEG signals during the passive listening task were recorded using a 64-channel EEG system (Biosemi Active 2 system, Biosemi Co., Netherlands) at a sampling rate of 2,048 Hz and band-pass (1-57 Hz) filtered. EEG signals were analyzed using a 3 s epoch length starting from -0.2 s prior to stimulus onset. The estimated potential was computed by averaging 100 epochs and filtered through a 1-15 Hz band-pass filter (butter worth order 5).

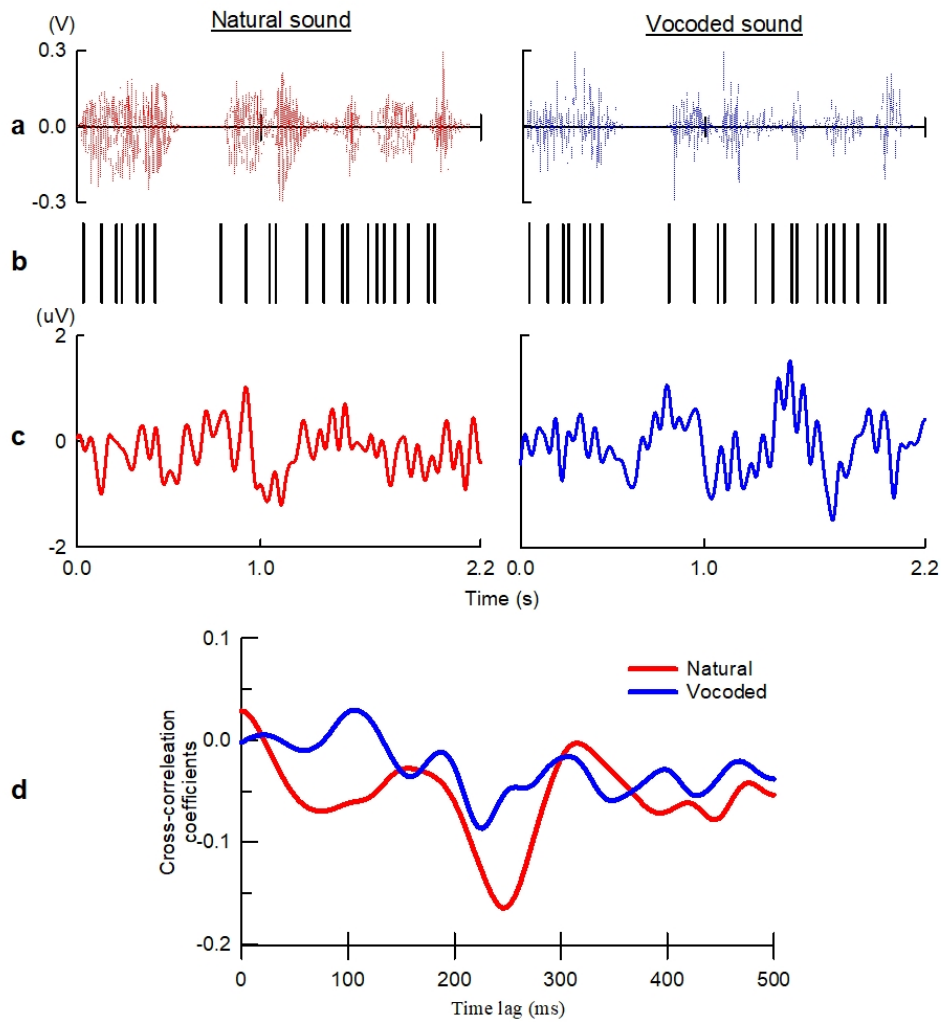
#### 2.1.5 Continuous speech-evoked potentials

**Figure 9** shows an example of (a) speech stimulation, (b) phoneme impulse train, (c) EEG, and (d) continuous speech-evoked potential (CSEP). The phoneme onset impulse and onset time of each phoneme in clean sentences and vocoded sentences were manually identified using the Praat program (Boersma, 2001). The cross-correlation coefficient was computed between a phoneme-onset impulse train and the evoked response potential. CSEP computed time lags between -1 and 1 s:

$$\text{CSEP}(ch, t) = \sum_T ph(T) \cdot EEG(ch, T + t)$$



where  $ph(T)$  and  $EEG(T)$  denote the phoneme onset train and the corresponding EEG response at time  $T$  and channel  $ch$ , respectively, and  $t$  denotes the time lag between phoneme onset train and EEG signal. The time lag of cross-correlation coefficients ranged from 0 to 500 ms. Cross-correlation coefficients were normalized by the mean of the time lag duration. Cross-correlations were averaged across trials and subjects.



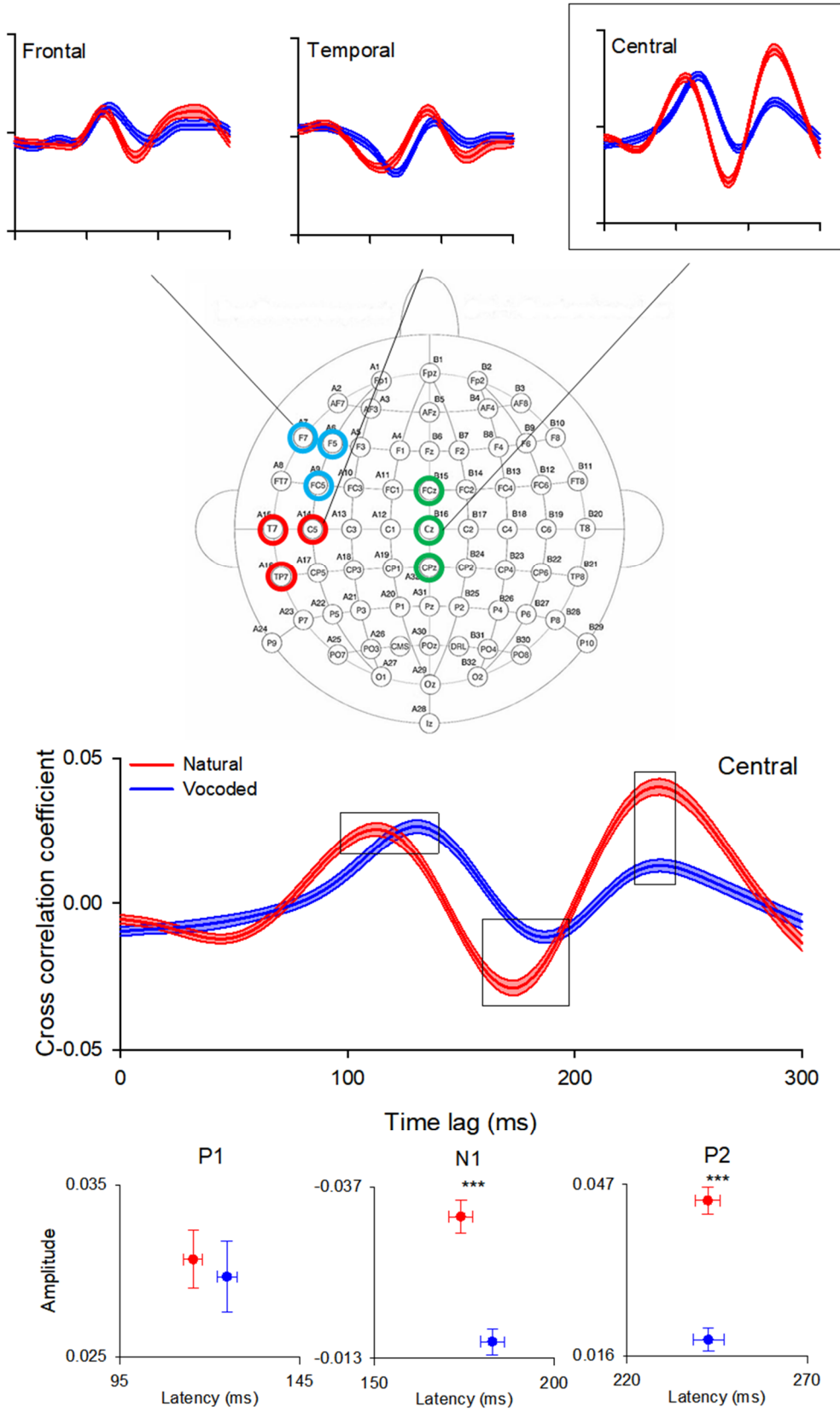
**Figure 9.** An example of CSEP extracting process between speech stimulus and EEG signal. **a.** Continuous speech wave form, **b.** phoneme onset impulse train, **c.** EEG signal evoked by continuous speech, **d.** phoneme onset neural tracking

The inverse problem was solved to calculate the source level of the evoked potentials through the MNE-python toolbox at the sensor level of 64-electrodes EEG signals. The phoneme-based CSEP of the source level was generated using cross-correlation between the source current intensity and the phoneme onset impulse train at each voxel. The normalization process was the same as the sensor level. The log p-values were calculated between the natural and vocoded conditions (Rank sum test, FDR correction) at each voxel.

## 2.2 Results

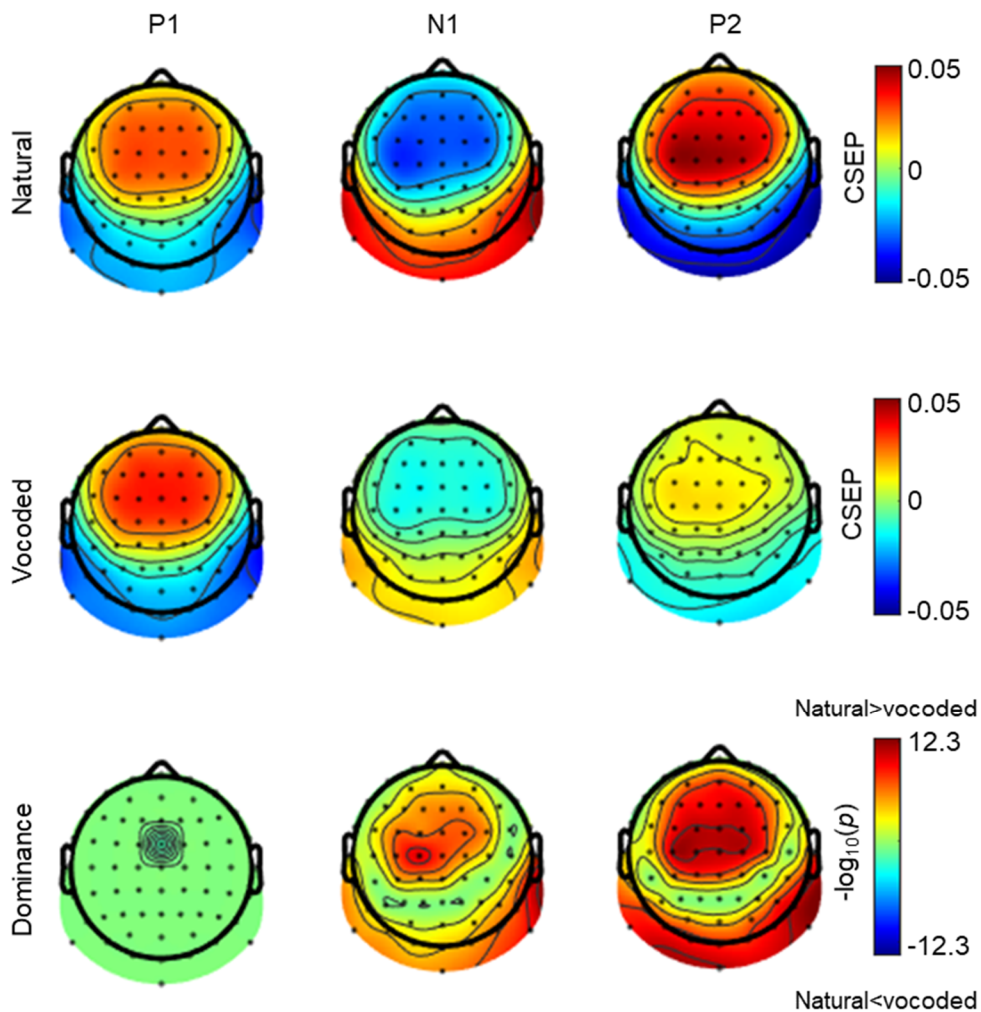
### 2.2.1 CSEP in sensor level

**Figure 10** shows an example of the grand phoneme-based CSEP from each listening condition within 0-300 ms. The morphology of phoneme-based CESP was comparable to typical auditory evoked potentials consisting of P1, N1, and P2 components.  $P1_{CSEP}$ ,  $N1_{CSEP}$ , and  $P2_{CSEP}$  amplitudes were compared between natural and 4-channel vocoded conditions. All three components ( $P1_{CSEP}$ ,  $N1_{CSEP}$ ,  $P2_{CSEP}$ ) observed at left-frontal (F7, F5, FC5), left-temporal (T7, C5, TP7), central regions. While the  $P1_{CSEP}$  amplitude within 90-150 ms at the central electrodes (FCz, Cz, Pz) had an insignificant difference between each condition, the N1 and P2 amplitude in the natural condition were significantly larger than that of vocoded conditions at the central electrodes (Wilcoxon signed-rank test,  $p < 0.001$ , FDR-corrected).



**Figure 10.** Grand averaged CSEP of natural (red) and vocoded condition (blue) at left-frontal, left-temporal, and central region (\*\*\*:  $p < 0.001$ , FDR-corrected).

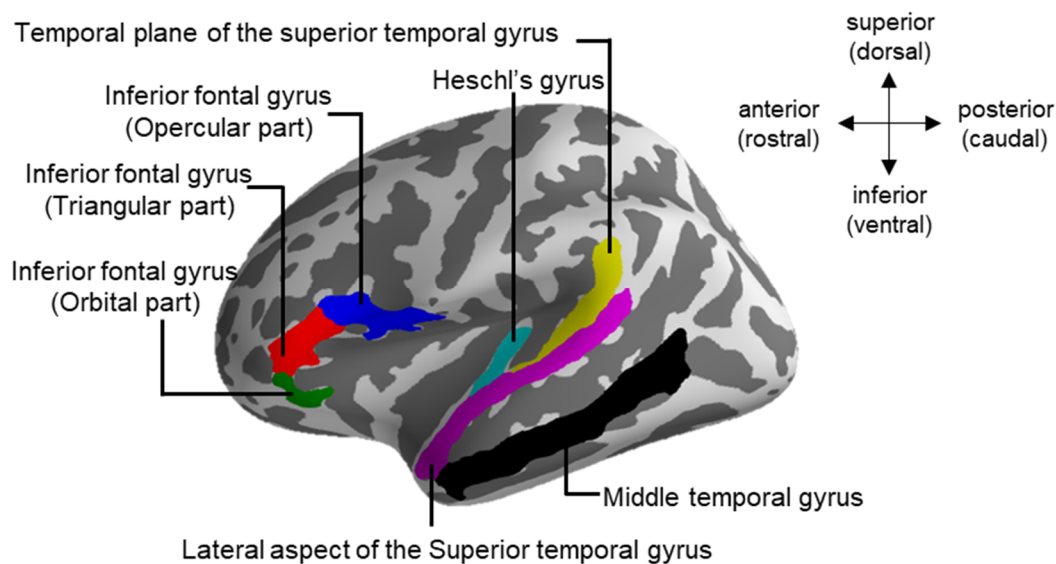
**Figure 11** shows the topographical maps of grand CSEP and the statistical difference between natural- and vocoded-conditions at  $P1_{CSEP}$ - $N1_{CSEP}$ - $P2_{CSEP}$  time lags. Topographical maps indicate differences in CSEP between natural- and vocoded conditions.  $N1_{CSEP}$  and  $P2_{CSEP}$  amplitude in the natural condition was significantly more extensive than that of the vocoded ones in the central areas (Wilcoxon signed-rank test,  $p < 0.001$ , FDR-corrected). Comparison of absolute CSEPs at significant electrodes showed (Wilcoxon signed-rank test,  $|\log(p)| > 1.3104$ ) a difference between natural and vocoded conditions. The dominant reaction may occur when the magnitude is more significant than that in the other conditions. In the case of CSEP, the natural condition was dominant in  $N1_{CSEP}$  and  $P2_{CSEP}$  at the central region.



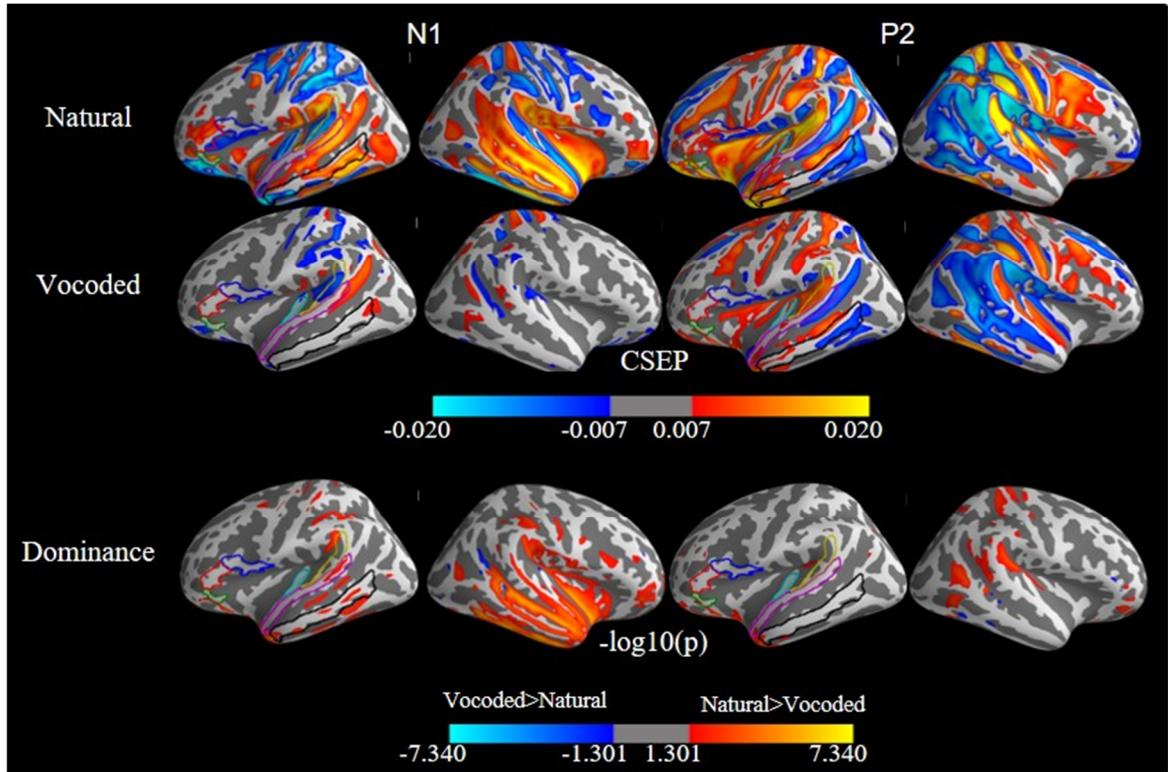
**Figure 11.** Grand averaged topographies at P1-N1-P2 complex latency.

### 2.2.2 CSEP at source level

Note that the source localizations of the two latter components ( $N1_{CSEP}$ ,  $P2_{CSEP}$ ) dominant well to the natural condition. **Figure 12** shows language processing-related areas in the left hemisphere. **Figure 13** shows the CSEP at source level at  $N1_{CSEP}$  and  $P2_{CSEP}$  latency and the dominance of both natural and vocoded conditions, with a color bar representing log p-values from 1.301 ( $p=0.05$ ) to 7.340 (Wilcoxon signed-rank test, FDR-corrected). As observed in **Figure 13**, natural condition dominant areas included the auditory cortex, inferior frontal gyrus, middle temporal superior gyrus, the temporal plane of the superior temporal gyrus. **Table II** and **Table III** show that the dominant condition was determined by the number of dominant voxels of each language processing area of **Figure 12** at  $N1_{CSEP}$  and  $P2_{CSEP}$ , respectively. The  $N1_{CSEP}$  of the natural condition is stronger than of vocoded condition at the middle temporal gyrus (MTG), lateral superior temporal gyrus (LSTG), the temporal plane of the superior temporal gyrus (STG), and orbital part of the inferior frontal gyrus (IFG). At the  $P2_{CSEP}$  response dominants to the natural condition at LSTG, the triangular part of IFG, and the orbital part of IFG.



**Figure 12.** Auditory language processing areas at source level.



**Figure 13.** Grand averaged CSEP in source level and dominance of voxels at N1-P2 complex latency

**Table II.** Dominance to conditions each language area at N1<sub>CSEP</sub>

Area (left hemisphere)	Natural dominant voxels	Vocoded dominant voxels	Dominant condition
Middle temporal gyrus	28.4% (52/183)	0.5% (1/183)	Natural
Lateral aspect of the superior temporal gyrus	19.3% (33/171)	0.6% (1/171)	Natural
Temporal plane of the superior temporal gyrus	10.9% (10/92)	1.1% (1/92)	Natural
Inferior frontal gyrus (Orbital part)	10.3% (3/29)	0% (0/29)	Natural
Inferior frontal gyrus (Triangular part)	-	-	-
Inferior frontal gyrus (Opercular part)	-	-	-
Heschl's gyrus (right hemisphere)	40.0% (14/35)	0% (0/35)	Natural

**Table III.** Dominance to conditions each language area at P2<sub>CSEP</sub>

<b>Area (left hemisphere)</b>	<b>Natural dominant voxels</b>	<b>Vocoded dominant voxels</b>	<b>Dominant condition</b>
<b>Lateral aspect of the superior temporal gyrus</b>	6.4% (11/171)	0.6% (1/171)	Natural
<b>Inferior frontal gyrus (Triangular part)</b>	4.9% (3/61)	0% (0/61)	Natural
<b>Inferior frontal gyrus (Orbital part)</b>	10.3% (3/29)	0% (0/29)	Natural
<b>Inferior frontal gyrus (Opercular part)</b>	-	-	-
<b>Middle temporal gyrus</b>	-	-	-
<b>Temporal plane of the superior temporal gyrus</b>	-	-	-
<b>Heschl's gyrus</b>	-	-	-

### 2.3 Conclusion & Discussion

In the present study, we proposed an objective approach to predict SI based on phoneme onset. The CSEP of a natural sentence follows phoneme onsets, and our results also show that CSEP is sensitive to SI.

A phoneme is a sound unit that distinguishes one word from another in a specific language (Reddy and Sajjan, 2021). Phoneme-level processing abstracts speech contrast differences, the basis of speech perception (Hessler et al., 2013). In a previous study, EEG signals revealed the response differences between phonemes in a sentence and that phoneme-related potential could be extracted from continuous speech-evoked EEG signals (Khalighinejad et al., 2017). In addition, the phoneme onset, which is one of the auditory transients, activates ERP even in passive listening conditions (Weise et al., 2012). The CSEP between different SI conditions significantly differed during passive listening in this study. Our results show that phoneme onset time represents the degree of speech intelligibility.

The IFG is known to be related to language production and comprehension processes. N1 represents the left prefrontal area and plays a role in language processing (Morin and Michaud, 2007; Paulesu et al., 1997; Poldrack and Wagner, 2004). N1<sub>CSEP</sub> was observed in the left frontal area dominant to natural conditions at the source level (**Table II**). Moreover, the natural sentences produced stronger responses of P2<sub>CSEP</sub> at LSTG, IFG than vocoded sentences (**Table III**). In other words, N1<sub>CSEP</sub> and P2<sub>CSEP</sub> responses represent SI in a passive listening task. In addition, the response to natural conditions is dominant in phonological or sentence-level speech processing areas such as the angular gyrus (Bonner et al., 2013), MTG (Bonner et al., 2013; Hickok and Poeppel, 2007), the temporal plane of the STG, and inferior circular sulcus of the insula (Oh et al., 2014) and Heschl's gyrus (Arsenault and Buchsbaum, 2015). The left temporal regions play a role in phonemic perception (Liebenthal et al., 2005; Scott et al., 2000).

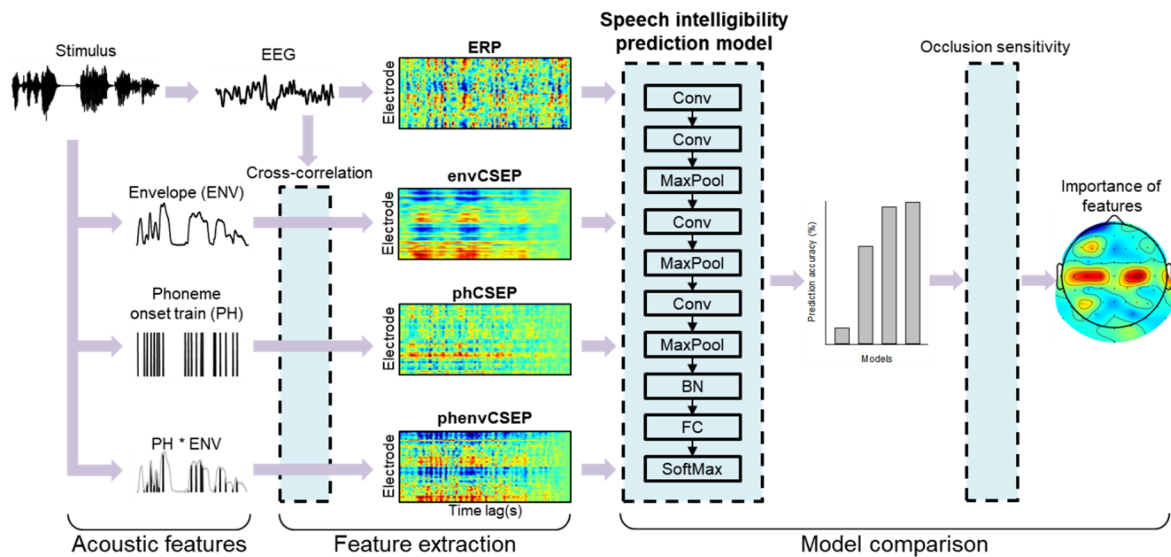
In conclusion, this Chapter II developed a novel approach for extracting continuous sentence-evoked EEG signals. Our results also showed that phoneme-based CSEP could evaluate speech intelligibility.



## **Chapter III. Predicting speech intelligibility using deep learning model**

Despite the phoneme-based CSEP is developed in Chapter II reveals a difference corresponding to speech intelligibility, it is still challenging area to quantitatively predict speech intelligibility scores from EEG signals to continuous speech. In addition, the phoneme information is not only important to comprehend continuous speech, but also the temporal envelope of speech. Here, we developed a novel objective approach to predict speech intelligibility scores based on a deep learning model with speech features of temporal envelope and phoneme information in Chapter III. **Figure 14** shows the flow chart of the speech prediction model. The input of the speech intelligibility prediction model is an electrode-time lag image of CSEP or ERP.

Occlusion analysis has been widely used in image classification to show the sensitivity of a pre-trained CNN to different areas of an input image (Zeiler and Fergus, 2014). The occlusion analysis can estimate which area of the image is the most essential for the classification. In this study, occlusion sensitivity was employed to interpret the importance of electrode each the features.



**Figure 14.** Flow chart of speech intelligibility prediction model from speech signal

### 3.1 Methods

#### 3.1.1 Behavioral test

EEG signals were recorded using a 64-channel EEG system from 87 Korean individuals with normal hearing. The participants listened to 10 Korean sentences and their corresponding degraded sentences in a passive listening mode. The sentences were spectrally 2-, 3-, 4-, 5-, and 8-channel noise-vocoded for the degraded conditions, and each sentence was repeated 100 times. The Korean sentence recognition test with natural, 2-, 3-, 4-, 5-, and 8-channel noise-vocoded sentences were randomly conducted to obtain the behavioral speech intelligibility scores prior to EEG data acquisition. The behavioral speech intelligibility score was evaluated in the Korean sentence recognition test using 41 discrete scores, defined at every 2.5 % between 0 and 100 %.

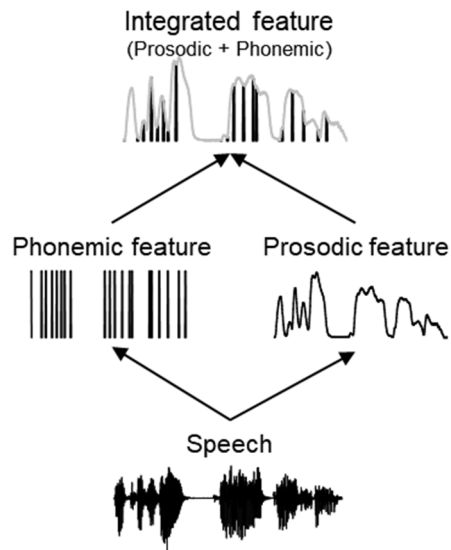
#### 3.1.2 EEG recording and procedure

EEG signals during the passive listening task were recorded using a 64-channel EEG system (Biosemi Active 2 system, Biosemi Co., Netherlands) at a sampling rate of 2,048 Hz and band-pass (1-57 Hz) filtered. EEG signals were analyzed using a 3 s epoch length starting from -0.5 s prior to stimulus onset. The EEG trials were randomly split into a training set, consisting of 80 % of the trials, and a test set, consisting of the remaining 20 % of the trials. Using bootstrap sampling, ERPs were computed by averaging 80 epochs in training set across each behavioral speech intelligibility score. The ERP filtered through a 1-15 Hz band-pass filter (butter worth order 5).

#### 3.1.3 Continuous speech-evoked potential with speech feature

The phoneme onset impulse train was identified using Praat software and manually confirmed. In this study, the phoneme onset impulse train, the speech envelope, and their convolution were used as speech features (**Figure 15**). The cross-correlation coefficient was computed between each speech feature and a single-trial EEG signal. The cross-correlation coefficients were averaged as a

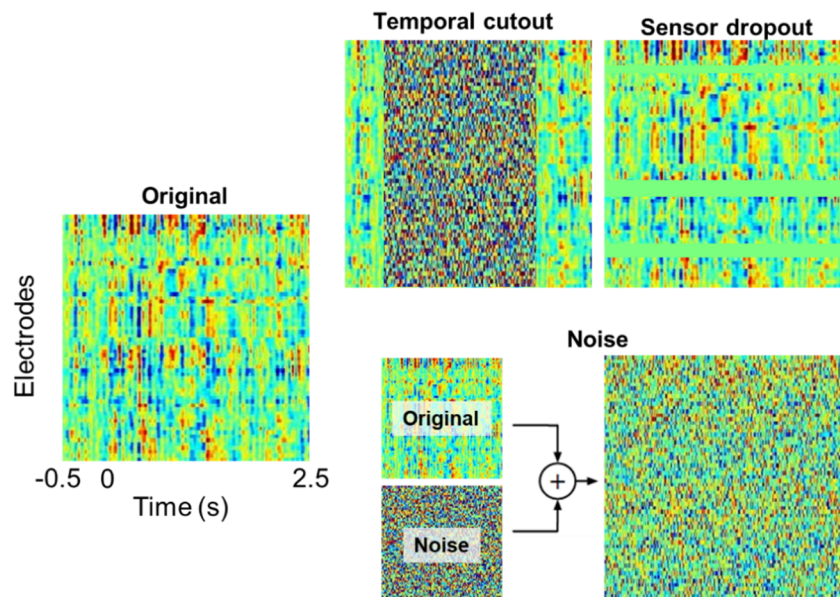
continuous speech-evoked potential (CSEP) across each behavioral speech intelligibility score.



**Figure 15.** An example of speech features from speech

#### 3.1.4 Data augmentation

The number of datasets in each class was imbalanced. To solve the imbalance problem and enlarge the number of training dataset, the training datasets were augmented with Gaussian noise, temporal cutout, and dropping sensors to guarantee the number of training datasets up to 8,000 (**Figure 16**) (Cheng et al., 2020).



**Figure 16.** Example of data augmentation.

### 3.1.5 Architecture of speech intelligibility prediction model

A deep learning model was trained using the features of ERPs, envelope-based CSEPs (ENV), phoneme-based CSEPs (PH), or phoneme-envelope-based CSEPs (PHENV) with the output of behavioral speech intelligibility scores. The architecture of the deep learning model consisted of four convolutional layers and a fully connected layer (**Table IV**). The convolution part employed max pooling, leakyReLU, and a batch normalization layer. The fully connected layer used the softmax layer for the activation function.

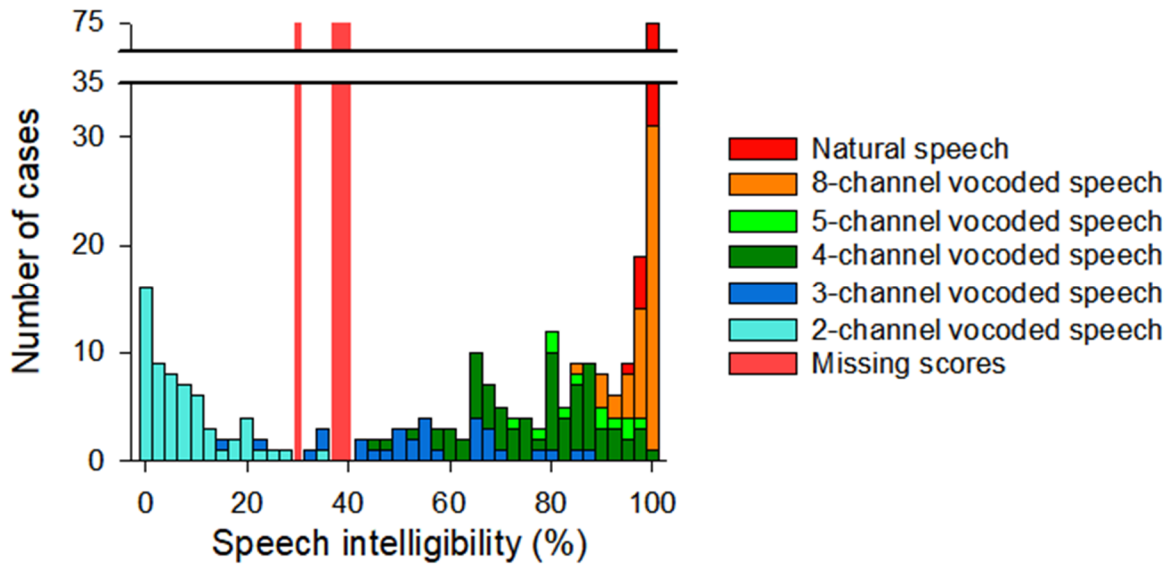
**Table IV.** Architecture of speech intelligibility prediction model

Type	Filters	Kernel	Output shape
<b>Input</b>	-	-	299 x 299 x 3
<b>Conv2D</b>	32	16 x 16	299 x 299 x 32
<b>LeakyReLU</b>	-	-	299 x 299 x 32
<b>Conv2D</b>	8	8 x 8	299 x 299 x 8
<b>LeakyReLU</b>	-	-	299 x 299 x 8
<b>MaxPooling2D</b>	-	2 x 2	149 x 149 x 8
<b>Conv2D</b>	8	4 x 4	149 x 149 x 8
<b>LeakyReLU</b>	-	-	149 x 149 x 8
<b>MaxPooling2D</b>	-	2 x 2	148 x 148 x 8
<b>Conv2D</b>	3	3 x 3	148 x 148 x 3
<b>LeakyReLU</b>	-	-	148 x 148 x 3
<b>MaxPooling2D</b>	-	2 x 2	147 x 147 x 3
<b>Batch Normalization</b>	-	-	147 x 147 x 3
<b>Fully Connected</b>	-	1 x 38	1 x 1 x 38
<b>Softmax</b>	-	-	1 x 1 x 38
<b>Classification</b>	-	-	38

## 3.2 Results

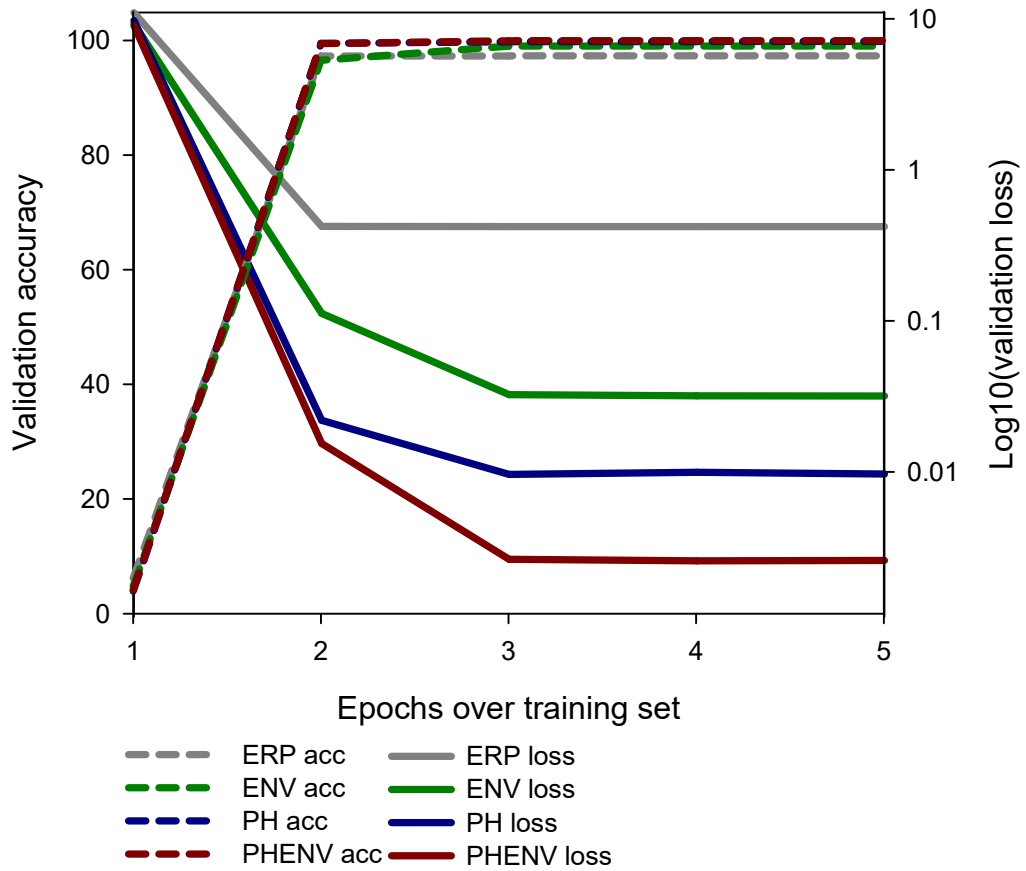
### 3.2.1 Speech intelligibility prediction model performance

The behavioral speech intelligibility scores of natural, 2-, 3-, 4-, 5-, and 8-channel noise-vocoded sentences covers the 38 ranges among 41 levels (**Figure 17**). The missing SI scores are 35.0, 42.5, 45.0%.

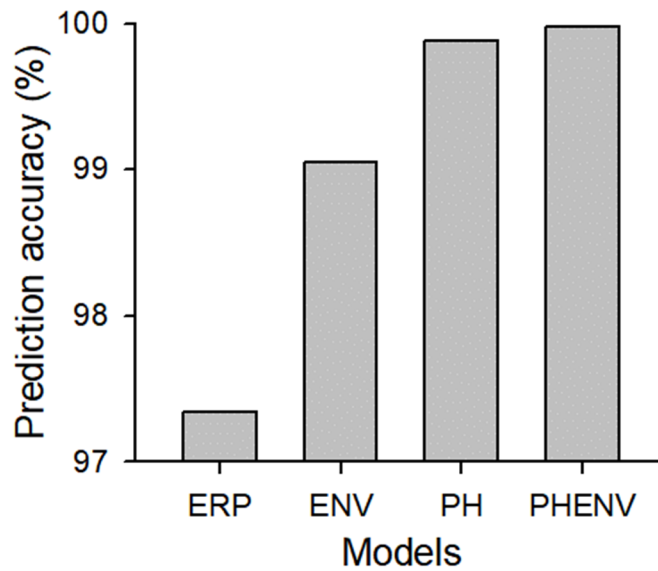


**Figure 17.** Summary of behavioral SI results. The bar color denoted by stimulus type.

During the first three training epochs, the validation loss of the deep learning model with the feature of ERP, ENV, PH, and PHENV decreased below 0.420 (training loss: 0.421), 0.032 (0.112), 0.010 (0.022), and 0.003 (0.015), respectively, and saturated thereafter. The deep learning models resulted in the good fit to the training data with no overfitting (**Figure 18**). The deep learning model predicted speech intelligibility scores with the test accuracy rates of 97.34 (ERP), 99.05 (ENV), 99.87 (PH), and 99.97 % (PHENV), which are comparable to the random chance level of 2.63 % (**Figure 19**).



**Figure 18.** loglized validation loss and accuracy across speech intelligibility models.

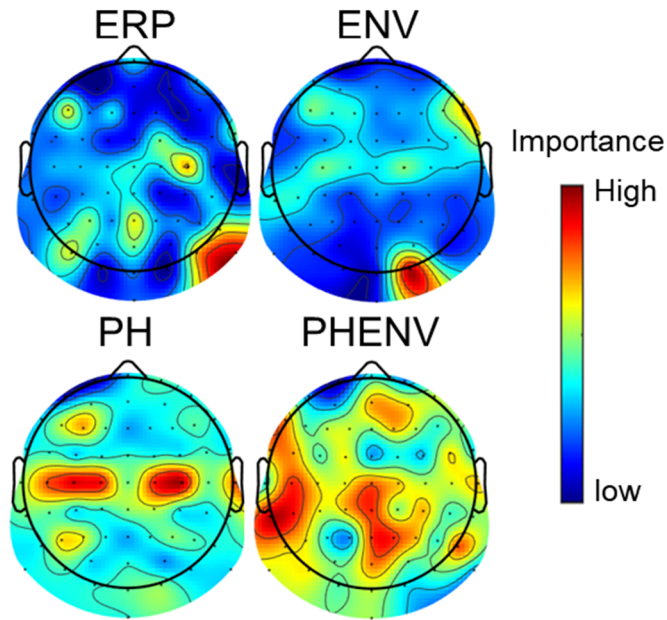


**Figure 19.** The performance of each speech intelligibility prediction model.

### 3.2.2 Occlusion sensitivity map

While the most important electrodes of the occlusion sensitivity map are typically right occipital in the ERP and ENV models, the electrodes in the PH and the PHENV model widely placed the left frontal, central, temporal electrodes are related with language processing are important for predicting SI scores (**Figure 20**).

**Table V** is a summary of top-10 informative electrodes to predict speech intelligibility.



**Figure 20.** Occlusion sensitivity topographies, color denoted by importance for speech intelligibility prediction

**Table V.** Informative electrodes for speech intelligibility prediction based on occlusion sensitivity maps

Feature type	Top-10 informative electrodes
ERP	P10, C4, PO8, Pz, P8, F5, PO7, P3, FC2, POZ
ENV	O2, F8, FT8, C2, F6, C3, F5, TP7, F1, C1
PH	C4, C3, C1, C2, C5, T8, P3, P5, P3, F5
PHENV	TP7, CP5, C5, CPz, P2, Pz, P8, Cz ,POz, F7



### 3.3 Conclusion & Discussion

This study developed a deep learning model to quantitatively and objectively predict speech intelligibility scores using continuous speech-evoked EEG signals. The results demonstrated that the deep learning model with EEG signals to continuous speech could accurately assess speech intelligibility. Moreover, the convolution of phoneme information and envelope of continuous speech was a more reliable feature for the deep learning model.

The occlusion sensitivity maps of PH and PHENV showed that the CSEP from the language dominant area resulted in better performance in predicting speech intelligibility (

**Table V**).

Davis and Johnsrude (2003) showed that brain areas were identified in which activation increased as intelligibility decreased; a left-lateralized frontal and temporal lobe system showed this profile. The left inferior frontal gyrus (IFG), including Broca's area, seems to be involved in processing complex auditory stimuli. The data support a role for the left IFG in syntactic and phonological processing (Friederici et al., 2000; Heim et al., 2003). The top-10 informative electrodes of both PH and PHENV include the electrode around the left IFG and temporal lobe (**Table V**). Thus, PH and PHENV could be reliable features to predict SI. Furthermore, the speech intelligibility model could be optimized using the occlusion sensitivity map, such as reducing the number of electrodes in future work.

The EEG-based deep learning model can be used as a valuable tool to objectively assess the benefit of the auditory prostheses and optimize them for better speech understanding in the near future.

## **Chapter VI. General conclusion**

The main purpose of this thesis is to develop an objective EEG-based speech intelligibility prediction model. Chapter II developed the phoneme onset CSEP to predict speech intelligibility and validated comparing within-subject. The results showed that the phoneme-based CSEP significantly differed between natural and vocoded conditions in the language processing area. Thus, the phoneme-based CSEP could predict the SI. The SI prediction model was developed in Chapter III. Comparing the performance of the SI prediction model, the PHENV is the best SI prediction model.

Furthermore, the informative electrodes of the PHENV model were related to the language processing area. Thus, the PHENV feature was most reliable for predicting behavioral SI scores from EEG signals. This dissertation's findings could prove that accurate SI prediction could be possible without listening fatigue and behavioral response.

Translating this method to the clinic needs to be further validated with a more diverse population with a broader age range, including children, in different languages. Furthermore, the occlusion sensitivity map is necessary to reduce the number of electrodes to ease the use of the SI model in a clinic in future work. The speech intelligibility prediction model should be capable of evaluating the efficacy of auditory prostheses (e.g., hearing aids, cochlear implants) in individuals with hearing impairments. However, auditory prostheses are electronic devices that generate power noise during operation. Thus, the artifact removing process is essential to use the SI model for hearing impaired patients in the clinic.

## Reference

- Accou, B., Monesi, M. J. and Francart, T.: Predicting speech intelligibility from EEG in a non-linear classification paradigm, *J. Neural Eng.*, 2021.
- Ag, P., Phonak, G. and Ag, C.: Speech intelligibility in dual task with hearing aids and adaptive digital wireless microphone technology, *Proc. Int. Symp. Audit. Audiol. Res.*, 6(August), 383–390, 2017.
- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H. and Merzenich, M. M.: Speech comprehension is correlated with temporal response patterns recorded from auditory cortex, *Proc. Natl. Acad. Sci.*, 98(23), 13367–13372, doi:10.1073/PNAS.201400998, 2001.
- Aiken, S. J. and Picton, T. W.: Human Cortical Responses to the Speech Envelope, *Ear Hear.*, 29(2) [online] Available from: [https://journals.lww.com/ear-hearing/Fulltext/2008/04000/Human\\_Cortical\\_Responses\\_to\\_the\\_Speech\\_Envelope.1.aspx](https://journals.lww.com/ear-hearing/Fulltext/2008/04000/Human_Cortical_Responses_to_the_Speech_Envelope.1.aspx), 2008.
- Arsenault, J. S. and Buchsbaum, B. R.: Distributed Neural Representations of Phonological Features during Speech Perception, *J. Neurosci.*, 35(2), 634–642, doi:10.1523/jneurosci.2454-14.2015, 2015.
- Boersma, P.: Praat, a system for doing phonetics by computer, *Glott. Int.*, 5(9), 341–345, 2001.
- Bonner, M. F., Peelle, J. E., Cook, P. A. and Grossman, M.: Heteromodal conceptual processing in the angular gyrus, *Neuroimage*, 71, 175–186, doi:10.1016/j.neuroimage.2013.01.006, 2013.
- Brodbeck, C. and Simon, J. Z.: Continuous speech processing, *Curr. Opin. Physiol.*, 18, 25–31, doi:<https://doi.org/10.1016/j.cophys.2020.07.014>, 2020.
- Brodbeck, C., Hong, L. E. and Simon, J. Z.: Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech, *Curr. Biol.*, 28(24), 3976-3983.e5, doi:<https://doi.org/10.1016/j.cub.2018.10.042>, 2018.
- Chen, S., Yang, E. Y., Kwan, M., Chang, P., Shiao, A. and Lien, C.: Infant hearing screening with an automated auditory brainstem response screener and the auditory brainstem response, *Acta Paediatr.*, 85(1), 14–18, 1996.
- Cheng, J. Y., Goh, H., Dogrusoz, K., Tuzel, O. and Azemi, E.: Subject-aware contrastive learning for biosignals, *arXiv Prepr. arXiv2007.04871*, 2020.
- Crosse, M. J., Di Liberto, G. M., Bednar, A. and Lalor, E. C.: The Multivariate Temporal Response

- Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli, *Front. Hum. Neurosci.*, 10(November), 1–14, doi:10.3389/fnhum.2016.00604, 2016.
- D., K. R., Gary, W., F., K. J. and C., R. J.: Toward Phonetic Intelligibility Testing in Dysarthria, *J. Speech Hear. Disord.*, 54(4), 482–499, doi:10.1044/jshd.5404.482, 1989.
- Das, N., Biesmans, W., Bertrand, A. and Francart, T.: The effect of head-related filtering and ear-specific decoding bias on auditory attention detection, *J. Neural Eng.*, 13(5), 56014, doi:10.1088/1741-2560/13/5/056014, 2016.
- Davis, M. H. and Johnsrude, I. S.: Hierarchical Processing in Spoken Language Comprehension, *J. Neurosci.*, 23(8), 3423 LP – 3431, doi:10.1523/JNEUROSCI.23-08-03423.2003, 2003.
- Decruy, L., Vanthornhout, J., Kuchinsky, S. E., Anderson, S., Simon, J. Z. and Francart, T.: Neural tracking of continuous speech is exaggerated in healthy aging and hearing impaired adults, 2021.
- Ding, N. and Simon, J. Z.: Emergence of neural encoding of auditory objects while listening to competing speakers, *Proc. Natl. Acad. Sci.*, 109(29), 11854 LP – 11859, doi:10.1073/pnas.1205381109, 2012.
- Ding, N. and Simon, J. Z.: Cortical entrainment to continuous speech: functional roles and interpretations, *Front. Hum. Neurosci.*, 8(May), 1–7, doi:10.3389/fnhum.2014.00311, 2014.
- Donhauser, P. W. and Baillet, S.: Two Distinct Neural Timescales for Predictive Speech Processing, *Neuron*, 105(2), 385-393.e9, doi:https://doi.org/10.1016/j.neuron.2019.10.019, 2020.
- Edwards, B.: The future of hearing aid technology, *Trends Amplif.*, 11(1), 31–45, 2007.
- Enderby, P.: Frenchay Dysarthria Assessment, *Br. J. Disord. Commun.*, 15(3), 165–173, doi:10.3109/13682828009112541, 1980.
- Friederici, A. D.: The brain basis of language processing: from structure to function, *Physiol. Rev.*, 91(4), 1357–1392, 2011.
- Friederici, A. D., Opitz, B. and Von Cramon, D. Y.: Segregating semantic and syntactic aspects of processing in the human brain: an fMRI investigation of different word types, *Cereb. cortex*, 10(7), 698–705, 2000.
- Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M. B., Dodel, R., Dubois, B., Holloway, R., Jankovic, J., Kulisevsky, J.,

- Lang, A. E., Lees, A., Leurgans, S., LeWitt, P. A., Nyenhuis, D., Olanow, C. W., Rascol, O., Schrag, A., Teresi, J. A., van Hilten, J. J. and LaPelle, N.: Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results, *Mov. Disord.*, 23(15), 2129–2170, doi:doi:10.1002/mds.22340, 2008.
- Healy, E. W., Yoho, S. E., Chen, J., Wang, Y. and Wang, D.: An algorithm to increase speech intelligibility for hearing-impaired listeners in novel segments of the same noise type, *J. Acoust. Soc. Am.*, 138(3), 1660–1669, doi:10.1121/1.4929493, 2015.
- Heim, S., Opitz, B. and Friederici, A. D.: Distributed cortical networks for syntax processing: Broca's area as the common denominator, *Brain Lang.*, 85(3), 402–408, 2003.
- Hessler, D., Jonkers, R., Stowe, L. and Bastiaanse, R.: Brain & Language The whole is more than the sum of its parts – Audiovisual processing of phonemes investigated with ERPs, , 124, 213–224, doi:10.1016/j.bandl.2012.12.006, 2013.
- Hickok, G. and Poeppel, D.: The cortical organization of speech processing, *Nat Rev Neurosci*, 8(5), 393–402, doi:nrn2113 [pii]r10.1038/nrn2113, 2007.
- Huang, Y. T.: Chapter 18 Sentence Processing., 2015.
- Hyde, M. L., Riko, K. and Malizia, K.: Audiometric accuracy of the click ABR in infants at risk for hearing loss., *J. Am. Acad. Audiol.*, 1(2), 59–66, 1990.
- Iotzov, I. and Parra, L. C.: EEG can predict speech intelligibility, *J. Neural Eng.*, 16(3), 36008, 2019.
- Jang, H., Lee, J., Lim, D., Lee, K., Jeon, A. and Jung, E.: Development of Korean Standard Sentence Lists for Sentence Recognition Tests, *Audiol*, 4(2), 161–177, doi:10.21848/audiol.2008.4.2.161, 2008.
- Karunathilake, I. M. D., Dunlap, J. L., Perera, J., Presacco, A., Decruy, L., Anderson, S., Kuchinsky, S. E. and Simon, J. Z.: Effects of Aging on the Cortical Representation of Continuous Speech, 2021.
- Khalighinejad, B., Cruzatto da Silva, G. and Mesgarani, N.: Dynamic Encoding of Acoustic Features in Neural Responses to Continuous Speech, *J. Neurosci.*, 37(8), 2176–2185, doi:10.1523/JNEUROSCI.2383-16.2017, 2017.
- Kim, G., Lu, Y., Hu, Y. and Loizou, P. C.: An algorithm that improves speech intelligibility in

noise for normal-hearing listeners, *J. Acoust. Soc. Am.*, 126(3), 1486–1494, doi:10.1121/1.3184603, 2009.

Kong, Y.-Y., Somarowthu, A. and Ding, N.: Effects of Spectral Degradation on Attentional Modulation of Cortical Auditory Responses to Continuous Speech, *J. Assoc. Res. Otolaryngol.*, 16(6), 783–796, doi:10.1007/s10162-015-0540-x, 2015.

Lee, J.: Standardization of Korean Speech Audiometry, , 9–11, 2016.

Di Liberto, G. M., O’Sullivan, J. A. and Lalor, E. C.: Low-frequency cortical entrainment to speech reflects phoneme-level processing, *Curr. Biol.*, 25(19), 2457–2465, 2015.

Di Liberto, G. M., Lalor, E. C. and Millman, R. E.: Causal cortical dynamics of a predictive enhancement of speech intelligibility, *Neuroimage*, 166, 247–258, doi:10.1016/J.NEUROIMAGE.2017.10.066, 2018.

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T. and Medler, D. A.: Neural Substrates of Phonemic Perception, , 15(10), 1621–1631, 2005.

Lightfoot, G.: Summary of the N1-P2 cortical auditory evoked potential to estimate the auditory threshold in adults, in *Seminars in hearing*, vol. 37, pp. 1–8, Thieme Medical Publishers., 2016.

Morin, A. and Michaud, J.: Self-awareness and the left inferior frontal gyrus : Inner speech use during self-related processing, , 74, 387–396, doi:10.1016/j.brainresbull.2007.06.013, 2007.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J. and Luuk, A.: Language-specific phoneme representations revealed by electric and magnetic brain responses, *Nature*, 385(6615), 432–434, 1997.

Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., Howard, M. A. and Brugge, J. F.: Temporal Envelope of Time-Compressed Speech Represented in the Human Auditory Cortex, *J. Neurosci.*, 29(49), 15564 LP – 15574, doi:10.1523/JNEUROSCI.3065-09.2009, 2009.

O’Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A. and Lalor, E. C.: Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG, *Cereb. Cortex*, 25(7), 1697–1706 [online] Available from: <http://dx.doi.org/10.1093/cercor/bht355>, 2015.

Oh, A., Duerden, E. G. and Pang, E. W.: The role of the insula in speech and language processing,

- Brain Lang., 135, 96–103, doi:10.1016/j.bandl.2014.06.003, 2014.
- Paulesu, E., Goldacre, B., Scifo, P., Cappa, S. F., Gilardi, M. C., Castiglioni, I., Perani, D. and Ca, F. F.: Functional heterogeneity of left inferior frontal cortex as revealed by fMRI, , 8(8), 2011–2016, 1997.
- Picton, T.: Hearing in time: evoked potential studies of temporal processing, Ear Hear., 34(4), 385–401, 2013.
- Poldrack, R. A. and Wagner, A. D.: What Can Neuroimaging Tell Us About the Mind ? Insights From Prefrontal Cortex, Curr. Dir. Psychol. Sci., 13(5), 177–181, 2004.
- Reddy, M. V and Sajjan, M.: Phoneme and Phone Pre-processing Using CLIR Techniques, in Data Science and Security, pp. 299–304, Springer., 2021.
- Robertson, S. J.: Robertson Dysarthria Profile, Buckinghamsh. Winslow, 1982.
- Sanders, L. D. and Neville, H. J.: An ERP study of continuous speech processing, Cogn. Brain Res., 15(3), 228–240, doi:10.1016/S0926-6410(02)00195-7, 2003.
- Scott, S. K., Blank, C. C., Rosen, S. and Wise, R. J. S.: Identification of a pathway for intelligible speech in the left temporal lobe, Brain, 123(12), 2400–2406, 2000.
- Shannon, C. E.: A mathematical theory of communication, Bell Syst. Tech. J., 27(3), 379–423, doi:10.1002/j.1538-7305.1948.tb01338.x, 1948.
- Stenfelt, S. and Rönnberg, J.: The Signal-Cognition interface: Interactions between degraded auditory signals and cognitive processes, Scand. J. Psychol., 50(5), 385–393, 2009.
- Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z. and Francart, T.: Speech intelligibility predicted from neural entrainment of the speech envelope, J. Assoc. Res. Otolaryngol., 19(2), 181–191, 2018.
- Venetjoki, N., Kaarlela-Tuomaala, A., Keskinen, E. and Hongisto, V.: The effect of speech and speech intelligibility on task performance, Ergonomics, 49(11), 1068–1091, doi:10.1080/00140130600679142, 2006.
- Weise, A., Bendixen, A., Müller, D. and Schröger, E.: Which kind of transition is important for sound representation? An event-related potential study, Brain Res., 1464, 30–42, doi:10.1016/J.BRAINRES.2012.04.046, 2012.



Weissbart, H., Kandylaki, K. D. and Reichenbach, T.: Cortical Tracking of Surprisal during Continuous Speech Comprehension, *J. Cogn. Neurosci.*, 32(1), 155–166, doi:10.1162/jocn\_a\_01467, 2020.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K. and Rabinowitz, W. M.: Better speech recognition with cochlear implants, *Nature*, 352(6332), 236, 1991.

Zeiler, M. D. and Fergus, R.: Visualizing and understanding convolutional networks, in *European conference on computer vision*, pp. 818–833, Springer., 2014.