Spring 2021

# How Does the Brain Predict Who's Speaking?

Olivia Sourwine

Inyong Choi
*University of Iowa*

HOW DOES THE BRAIN PREDICT WHO'S SPEAKING?

by

Olivia SourwineInyong Choi

A thesis submitted in partial fulfillment of the requirements
for graduation with Honors in the Speech Pathology and Audiology

Inyong Choi
Thesis Mentor

Spring 2021

All requirements for graduation with Honors in the
Speech Pathology and Audiology have been completed.

Yu-Hsiang Wu
Speech Pathology and Audiology Honors Advisor

College of Liberal Arts and Sciences
The University of Iowa
Iowa City, Iowa

# HOW DOES THE BRAIN PREDICT WHO'S SPEAKING?

By
Olivia Ann Sourwine

A thesis submitted in partial fulfillment of the requirements for graduation with Honors in the Department of Communication Sciences and Disorders

Dr. Inyong Choi
Thesis Mentor

Spring 2021

## Abstract

While sitting in a noisy environment you may have trouble understanding your conversation partner. However, listening to a familiar voice of a friend may be easier compared to listening to an unfamiliar voice. Past research studies support this phenomenon with evidence of stronger speech-evoked brain activity while listening to a familiar speaker. Additionally, previous studies have developed the concept of the Predictive Coding Theory. This theory states that the brain predicts what will occur in its sensory environment based on its internal representations of the world. The incoming sensory inputs update the predictions. There has been an emphasis on the physiological evidence of neurocognitive processing in this theory. One in which is dominated by top-down processing, and the other uses sensory sampling. In this study, we claim that gamma band oscillations use sensory sampling and beta band oscillations use top-down processing and predicting. However, there is no evidence on what neural substrates are involved while tracking a speaker's identity in noise. Our goals were to investigate the cortical oscillations produced and their location in the brain when there is a speaker identity cue versus no cue. We measured cortical EEG data of 13 normal-hearing participants in speech-in-noise. The speaker identity cues increased beta band oscillations in the inferior frontal gyrus region of the left hemisphere. However, without a speaker identity cue, greater gamma band oscillations were found in the supratemporal gyrus in both hemispheres. The results had significant differences amongst both conditions. With these results, we support the Predictive Coding Theory and the physiological evidence of cortical oscillations. While the brain tracks a familiar voice in noisy environments it uses predictive top-down processing indicated by beta waves. However, if the brain cannot predict the speaker, then sampling of auditory features occurs, indicated through gamma oscillations. This allows the brain to be open to all potential voices.

## Introduction

In an active world full of events, it allows for our surrounding environments to be rich in sound. High levels of background noise are a common trend in our everyday lives.  However, understanding speech in noise can be a difficult task for everyone. The background noise causes the communicative message of speech to compete with the noise to be heard by the listener. The background noise degrades the speech signal, which in return causes difficulty for the listener to decode the message that is transferred through the environment to the ear. The background noise causes an obstacle to our communication with others around us. This hindrance does not only impact normal-hearing listeners but individuals with hearing loss as well. Individuals with cochlear implants can understand speech in quiet environments but it becomes increasingly worse with background noise (Hochmair-Desoyer et al., 1997; Wilson & Dorman, 2008; Zeng et al., 2011). The speech sound is degraded while being transmitted into an electrical signal with their cochlear implant device (Drennan & Rubinstein, 2008). This contributes to the hearing-impaired population and their most frequent complaint of not being able to understand their conversation partner in noisy environments.

Past researchers found that cochlear implant users listening to speech in different background noise conditions had slower higher-order processing compared to normal-hearing listeners. Their speech understanding outcomes are due to the differences in their sensory and auditory-cognitive processing. Additionally, mapping their acoustic-phonetic features to their lexical representations in these conditions is a barrier (Finke, Büchner, Ruigendijk, Meyer, & Sandmann, 2016). Therefore, individuals with cochlear implants experience high levels of strain and fatigue while listening in noisy backgrounds (Ohlenforst et al., 2017). Being so speech in background noise is a frequent complaint of individuals with hearing loss, we hope to serve this population in the future. Although we hope to make clinical advancements in individuals with hearing loss, we will first be investigating normal-hearing listeners to understand the normal neural mechanisms. This knowledge will open doors to better understand and help hearing-impaired listeners.

Using selective attention while attending to a stimulus in complex acoustic environments is beneficial (Holmes, Kitterick, & Summerfield, 2018). While listening to speech in background noise you are using neural mechanism strategies to track the conversation we wish to attend to and will focus on that stimulus. In our experiment, they will track and attend to a cued speaker's voice. The listener decides to ignore the other voices. This is supported by past research in which participants were presented with a cued acoustic target stimulus before the target during a selective attention task. They performed better compared to participants that had no cue presented or when there was a simultaneous cue (Koch, Lawo, Fels, & Vorländer, 2011; Lu et al., 2009; Richards & Neff, 2004). Additionally, in Kahneman's work (Kahneman, 1973), subjects had higher accuracy scores when presented with a cue before each trial. In their study, the cue directed their attention to a designated visual object (Kahneman, 1973, chapter 5). Before the listener even hears the target sound, they prime their cortical representations to become biased towards what their attention is directed towards (Voisin et al., 2006).

Additionally, Holmes et al. (Holmes, Kitterick, & Summerfield, 2018) investigated the duration between the cue-target interval in a multi-talker listening task. Across different trials, they varied the length of time between the cue and the target. Results indicated that the participants had better accuracy, accurately reported target words, and had shorter reaction times with the 2000 ms cued-target interval when compared to the 0 ms interval. Therefore, intelligibility progressively improves as the duration between the cue and target increases. This provides evidence that there is underlying preparation that takes place as their attention unfolds over time. The listener can prepare for the target talker's gender or location (Holmes, Kitterick, & Summerfield, 2018). During speech-in-noise tasks, the cue should be presented before the target and have a 2000 ms or longer cued-target interval. This will allow the listener to prepare for the target. Therefore, recognizing the speaker's identity through cueing with an increased duration between the cue and the target can benefit the listener's speech perception in background noise. In our study, will be using this method supported by past research of increasing the duration time amongst our speaker identity cue and the target word. Additionally, the time frame between the cue and the target is where we will be investigating the cortical oscillations that are produced by the brain.

Another reason why individuals with hearing loss face challenges while listening to speech in noisy environments is due to the perceptual discontinuity of the scene. For example, the speaker may move around throughout the environment or the speaker's role may be passed to a different individual. Mehraei et al. (Mehraei, Shinn-Cunningham, & Dau, 2018) investigated how discontinuity in the speaker identity affects their ability to use auditory attention towards localized sounds. Additionally, they investigated the underlying processes of the neural correlates. Their participants were asked to listen to a stream of spoken syllables from a talker. However, on some trials, the talker switched locations in the middle of the stream which caused perceptual discontinuity. While there was discontinuity within the speaker it disrupted the attentional modulation of the cortical response, suppression of alpha power oscillations, and the event-related potentials amongst the syllables were impacted. Therefore, speaker identity continuity is important and influences listening and selective auditory attention in a multi-source environment (Mehraei, Shinn-Cunningham, & Dau, 2018). With the stimulus of a continuous target voice and being able to attend to an auditory object in a noisy environment, the attentional selectivity improves (Best, Ozmeral, Kopčo, & Shinn-Cunningham, 2008).

Lastly, a familiar voice/speaker identity has been found to improve speech perception in acoustically complex environments. This is the reason why you can easily pick out your friend's voice during a conversation in a noisy environment. Past research provides rich findings showing evidence in stronger speech-evoked brain activity when speech is produced by a familiar voice during competing speech. Holmes and Johnsrude's (Holmes & Johnsrude, 2020) experiment had the participant listen to a sentence spoken by a familiar voice (the participant's friend or partner) and an unfamiliar voice (a different participant's friend or partner). Some conditions had the target presented alone and others had the familiar and unfamiliar voice saying sentences at the same time. This created a competition between the different speech streams produced by different voices. They compared the multivariate activity in speech-sensitive regions of the cortex amongst the two different conditions. In their study, they utilized representational similarity analysis. Their results supported that when the familiar speaker said the competing sentence versus the unfamiliar speaker, the activity evoked of the spoken sentence was less degraded. Therefore, listening to a familiar voice benefits intelligibility. Additionally, it showed that nonprimary auditory cortical areas were most prominent when there was an advantage of using the familiar voice (Holmes & Johnsrude, 2020).

Past researchers have developed the concept of the Predictive Coding Theory. This theory is defined as, "the idea that the brain generates hypotheses about possible causes of forthcoming sensory events and that these hypotheses are compared with incoming sensory inputs, that is, the prediction error is propagated forward throughout the cortical hierarchy" (Arnal & Giraud, 2012). It provides a neurophysiological basis for predicting what will occur in the brain's sensory environment and the most probable cause of the repetitive stimuli (Arnal & Giraud, 2012; Lochmann & Deneve, 2011). Many consistent sensory regularities occur in our daily lives which allows the brain to host internalized representations of the world (Friston, 2005). If prediction errors occur, the brain is required to update its mental representations through sensory sampling.

There is physiological evidence that we use neurocognitive processing, one of which is dominated by top-down processing and the other uses sensory sampling. While listening to speech in background noise both bottom-up acoustic cues and top-down cognitive processes are incorporated (Strait & Kraus, 2011). This can be investigated through electroencephalogram (EEG) data. While the brain is predicting or using sensory sampling there are low-level oscillatory cortical mechanisms that have distinct neural rhythms based on the task (Arnal & Giraud, 2012). In our particular study, we will be focusing on gamma band oscillations and beta band oscillations. Gamma oscillations have a frequency greater than 30 Hz and function in involvement in various cognitive processing to encode and sample sensory stimuli. It is involved in attention, stimulus selection, and multisensory and sensorimotor integration (Donner & Siegel, 2011; Engel, Fries, & Singer, 2001; Fries, Nikolić, & Singer, 2007; Tallon-Baudry et al., 1996). The gamma band carries ascending information to the brain where it interprets sensory signals (Busch, Dubois, & VanRullen, 2009; Wang, 2010). Therefore, gamma band oscillations are dominated by bottom up-processing. Additionally, electrophysiological data shows that if a sensory stimulus is expected then the gamma band activity decreases (Arnal & Giraud, 2012; Todorovic et al., 2011; Gruber & Müller, 2005). This would also occur during the process of making predictions. In contrast, beta band oscillations have a frequency between 15-20 Hz and are associated with motor functions (Engel & Fries, 2010; Jenkinson & Brown, 2011). Research supports that beta carries descending information through top-down processing (Wang, 2010). The brain applies its prior knowledge to fill in the gaps and is heavily reliant on its predictions and expectations.

Arnal and Giraud’s work integrated the Predictive Coding Theory with the cortical oscillations and their computational solutions throughout the brain prediction process. Their data proposes that gamma activity is present during a sensory surprise and is utilized to signal unexpected occurrences and prediction errors. Additionally, their data supported that beta activity instead signals downstream processing to encode sensory inputs only if it is accurately predicted by the brain (Arnal & Giraud, 2012).

Not only do different types of cortical oscillations have distinct functions but also different regions of the brain specialize in their own functions. The supratemporal gyrus is located in the temporal lobe and the auditory cortex. These areas are associated with auditory sensory areas. The inferior frontal gyrus (Broca’s area) is located in the frontal lobe and is notorious for higher-level speech and language areas.

Currently, the literature is rich with studies exploring the Predictive Coding Theory exhibiting functions of cortical oscillations. However, less is known about the neural substrates that are involved while tracking a speaker’s identity in speech and noise. Therefore, this study aimed to investigate the cortical oscillations present and their locations in the brain when an auditory cue with a speaker identity was provided versus one without a speaker identity. In the cued condition you are predicting the future voice. However, with the no cue you have to be open for multiple voices. In this study, we used EEG to observe the neural substrates during the “preparatory period”. This is the time interval when the auditory cue was presented before the presentation of the target speech embedded in babble noise. We proposed that the preparatory period would be the period where the predicting or tracking of the speaker identity would take place. It was hypothesized that; the beta band oscillation would be present in the inferior frontal gyrus for the speaker identity cued condition. We also hypothesized that the gamma band oscillation will be located in the supratemporal gyrus for the condition with no speaker identity cue. We hope to better understand the neural mechanisms of cognitive processing during speaker identity tracking to make clinical advancements for individuals with hearing loss in noisy environments.

## Materials and Methods

### Participants

Thirteen participants with normal hearing participated in this study. (11 female and 2 male). The mean age was 21. The subjects were recruited within the University of Iowa’s student population. Before the experiment, all subjects voluntarily signed an informed consent form approved by the University of Iowa’s IRB. Then the IRB protocol was followed throughout the entire experiment. Subjects also filled out a demographic questionnaire and were compensated for their time.

The inclusionary criteria for the normal-hearing listeners were that they were young hearing adults between the age of 18-40 years old. They also must be fluent in English. English did not have to be their native language, but they needed to have strong English vocab and no spoken word perception issues. Additionally, after given a hearing screening their audiogram must have displayed thresholds below 25 dB for frequencies 250, 500, 1000, 2000, 4000, 6000, and 8000 Hz for both the left and right ear. These frequencies are where the speech sounds live and are perceived. Lastly, they were given a neurological questionnaire and could not have ADHD, a brain injury, report that they have been taking psychoactive drugs, or have other neurological conditions (Wickhman, Geller, & Choi, 2020).

### Stimuli

The stimuli of the experiment were real English words that originated from the Iowa Test of Consonant Perception (ITCP) (Geller, McMurray, Choi, & Holmes. A, 2020). The experiment consisted of a total of 120 balanced initial consonant and monosyllabic words. All words were consonant-vowel-consonant (CVC) structured. During the experiment, the keywords appeared in a closed set of four words that all rhymed. They only differed by the initial consonant of the word, differed by one phonetic feature, and were minimal pairs. The same words would always appear in the same group of four words every time. Each group of words cycled through multiple times, but the target word differed every time. The phoneme class and the number of trials for each phoneme were balanced. In ITCP there were two female voices and two male voices that were recorded for each English word. However, for this study, they only used one female voice, a male voice, and a mixture of both voices speaking at the same time. They ensured that the words spoken would all have the same duration, clear articulation, and prosody. All the voices were native English speakers and had standard American English accents.

Additionally, the target word was presented in multi-speaker babble noise (Kim et al., 2020; Geller, McMurray, Choi, & Holmes. A, 2020). It sounded as though the individual was sitting in a noisy restaurant with multiple conversations occurring in the background. The babble noise was created with a Matlab script. The babble noise originated from the spoken noise made by eight talkers and was five minutes long from the Revised SPIN Test (a compact disc published by Auditec, Inc., St. Louis, Missouri). However, in the experiment, the babble noise was cut down to 1.8 seconds while being played in the background. The babble noise varied across all trials. However, there was a limit on noise difference. The noise was always within 1.5 dB across all trials and the amplitude was consistent. The target word from the ITCP started after 2 seconds of the babble noise. The multiple-speaker babble noise would continue while the target word was presented. The fixed signal-to-noise ratio (SNR) for the subjects was 0 dB.

The equipment used for the experiment was an electrically shielded soundproof booth where the task took place. Additionally, the sound and stimuli were presented from a loudspeaker (model #LOFT40, JBL) with a 0-degree azimuth angle and position of 1.2 m. An amplifier was also used.  The presentation level was calibrated at 70 dB SPL. The presentation level equaled the noise plus the target word that was spoken. A wooden chair and table inside and outside of the sound booth were used to prevent and avoid metal-related artifacts in the EEG and Polhemus data. Furthermore, Matlab was used to implement the experiment on the computer monitor. The participant used the custom-made keypad designed with keys 1-4 in a horizontal line. The participant could easily map each finger to the four answer choices they saw on the computer monitor. The computer was positioned .5 meters and at eye level of all the subjects so they could see the answer choices.

## Experimental Design

During the task, the participant sat in the booth and used the computer monitor to complete the speech-in-noise task. The experiment had two different conditions. The first one was the cued condition in which the speaker identity was a single male voice or female voice. The voice was used in the carrier phrase to prime the target word later presented in the background noise with the same speaker identity. The cued condition would never have both speaker identities. However, there was also a non-cued condition. The ambiguous cue was when the male and female voice spoke at the same time during the carrier phrase. However, only one of the voices, either male or female, would say the target word presented later on (Wickham, Geller, & Choi, 2020). The subject was not primed with the speaker identity to know which gender was going to say the target word later.

There was a total of 240 total trials. After every 30 trials, there was a built-in optional break where the subject could relax, rest their eyes, or stretch. This would ensure the participant would not experience fatigue and keep their minds fresh. The task was a total of 240 trials so that all 120 words from ITCP each showed up twice. One time it was cued, and the second time it was non-cued. Out of the 120 words that were cued, 60 of them used the female speaker identity cue and the other 60 were the male's identity. Counterbalancing across all the participants was utilized for the cued condition (Wickham, Geller, & Choi, 2020). This would ensure that all 120 ITCP words could have been cued with the male speaker identity and also the female speaker identity. The counterbalancing across the subjects ensured there was an equal amount of both, and the same words appeared. It was randomized amongst subjects. At the end of each trial, there was no feedback on accuracy.

At the beginning of the task, there was a cross ('+') located at the center of the screen. The participants stared at the cross throughout the trial because it would eliminate artifacts of eye movement. The trial structure is represented in Figure 1 below. At the start of every trial, the subjects were asked to listen to the carrier phrase, "choose the word", which either was spoken by the speaker identity cue (single male or female voice) or by the ambiguous identity cue (both female and male voices at the same time). The carrier phrase would also prepare the subjects to listen for the target word presented in background noise. The carrier phrase was spoken 1.5 seconds before the onset of the multispeaker babble noise that began at 0 seconds. The target word that either matched or mismatched the auditory cue was presented in background noise at 1 second and lasted about .3-.4 seconds depending on the target word. The offset of noise and speech occurred at 2 seconds followed by silence.

Next, the screen would show a foiled list of four potential words that all differed by their initial consonants and were minimal pairs. The participants would then select the target word they thought they heard in the background noise by pressing the number on the keypad that corresponded to the word on the screen. Afterward, the next trial began immediately.

Additionally, if the subject took longer than 10 seconds to select the target word by pressing a key, then the next trial began and counted the trial as inaccurate. Accuracy was measured by the responses of the selected keys. If they correctly identified the target word in noise it was counted as correct. However, if they selected any of the three other options it was counted as incorrect. Accuracy was compared across both cued and non-cued conditions.

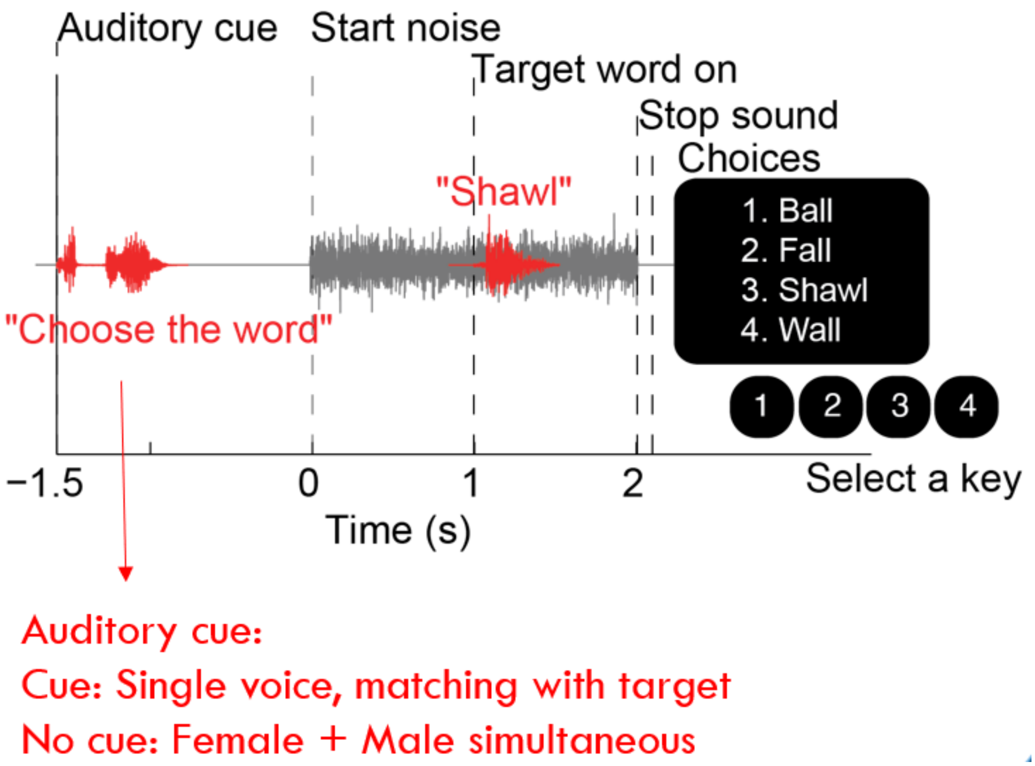

**Figure 1**: Trial structure of the auditory cue or ambiguous cue embedded in the carrier phrase, “choose the word”, before the onset of noise. Multi-speaker babble noise started at zero seconds and the target word either matched or mismatched the auditory cue that was presented in background noise. The offset of noise and speech was at 2 seconds. 4 minimal pair options appeared on the computer screen. The participant selected the key that corresponded to the target word they heard in the background noise.

## Electroencephalogram (EEG) Acquisition & Preprocessing

All the subjects wore a 64 channel EEG cap from the BioSemi ActiView system. After the cap was fit securely on their head, all the electrodes' positions were recorded and measured through the Polhemus Patriot 3D scanner system. Afterward, sixty-four active electrodes were inserted based on the international 10-20 configuration. The BioSemi ActiView system was utilized in the speech-in-noise task to record scalp electrical activity (EEG). The sampling rate was set to 2048 Hz. During the experiment, Matlab (R2016b, The Mathworks) sent the trigger signals to the BioSemi ActiView acquisition software. The offset voltages were maintained under 30mV.

Bandpass filtering was used for the EEG data. Every recorded channel was filtered from 1 to 20 Hz with a 2048-point zero-phase FIR filter. The epochs were extracted between -500 ms to 3 seconds relative to the noise onset. The average voltage between -200 and 0 ms relative to the target word onset was utilized for baseline correction. The epochs were down sampled to 256 Hz. Re-referencing was not done for the sensor-space analysis. However, re-referencing towards the all-channel average before the source-space analysis was used for EEG data (Kim et al., 2020).

Additionally, contaminated trials with eye blink artifacts were removed through independent component analysis. Their rejection was determined by the voltage value of the Fp1 electrode (bandpass filtered between 1 and 20 Hz). The range of rejection thresholds ranged from 35 to 120 uV (mean 62.7 and standard deviation 23.7) They were extracted through signal space projection. Then noisy epochs were rejected that passed 70 uV. After rejecting trials with artifacts, the electrode's averages across the mean number of trials were each calculated to extract event-related evoked potentials.

## Results

### EEG Analysis

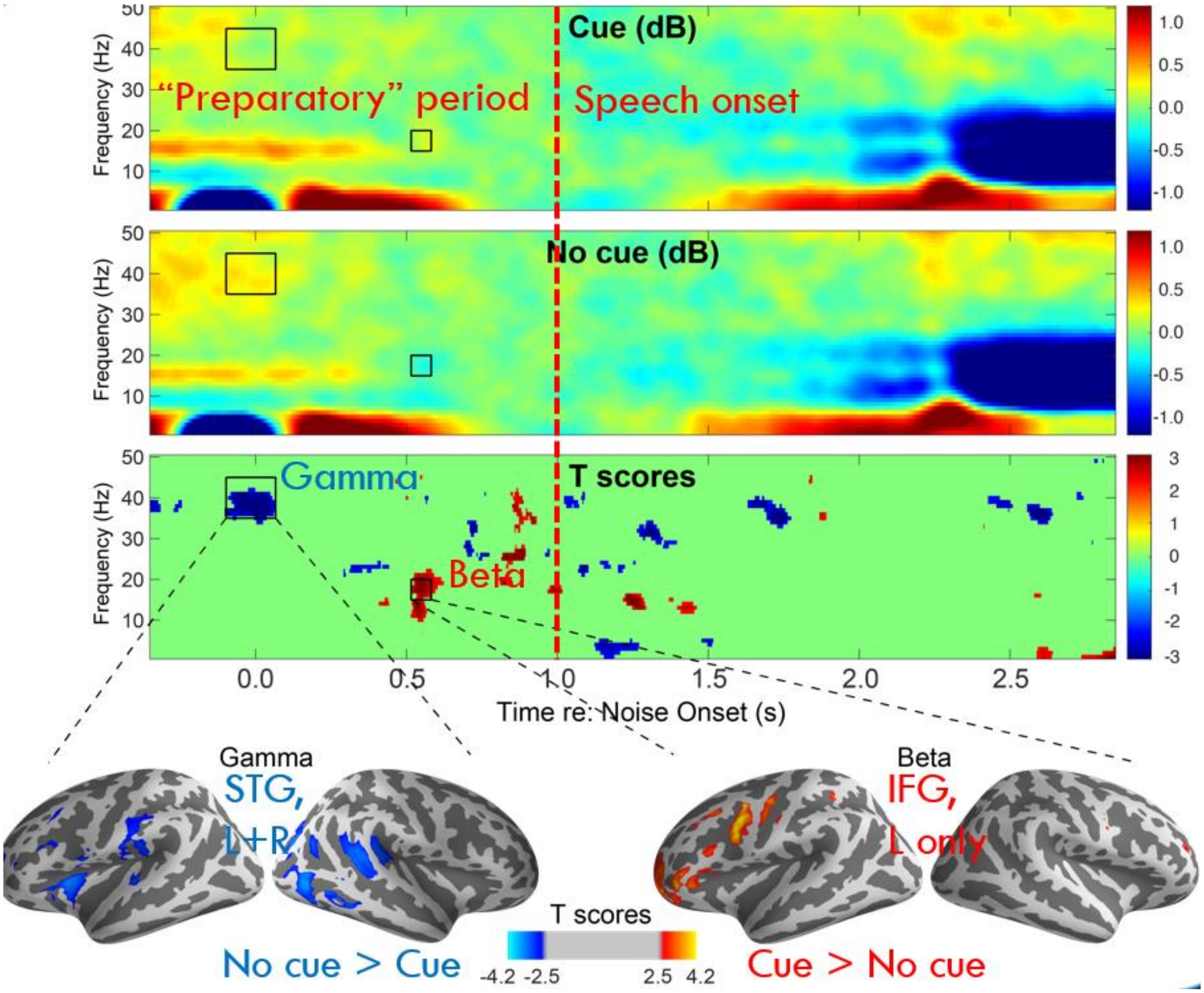

Figure 2: Spectrograms and cortical surface maps of induced activities. Black boxes around 0 and 0.5 seconds denote clusters of time-frequency bins that exhibited statistically significant differences between conditions. The first row: Grand-average spectrograms averaged across all 64 electrodes in the cue condition. The second row: Grand-average spectrograms averaged across all 64 electrodes in the no-cue condition. The between-condition t-scores in the third row are masked by p-value (<0.05, uncorrected).

The cortical surface map at the bottom shows the distribution of t-scores from paired t-tests performed within each cortical voxel between conditions within the significant clusters of the gamma (35 – 45 Hz, left panel) and beta (15-20 Hz, right panel) bands. A significant difference in gamma is found in the supratemporal gyrus in both hemispheres and beta was only found in the left hemisphere of the inferior frontal gyrus.

Analysis was conducted across the time course from -1.5 seconds after the auditory cue to 1 second before the speech target (preparatory period). The three panels are grand-average magnitude spectrograms with time on the x-axis, frequency on the y-axis, and intensity is based on the relative darkness of the colors (Figure 2). The spectrograms are averaged across all 64 channels. Analyses were done across the different oscillation bands for both the cued and non-cued conditions. Additionally, the bottom panels show t-scores from a series of paired t-tests across all the time-frequency bins. T-scores are used to show the difference in effect size of the cued and no cued condition. The red color represents positive values which indicate that the cued condition had statistically significant and greater effects than the no cued condition. In contrast, the blue color is negative and shows that the no cued condition had statistically significant and stronger effects. The neutral green color surrounding the blue and red colors indicates p-values > 0.05 (uncorrected) and there was no significant difference. Referencing back to the first panel in Figure 2 there is a strong frequency band of a beta oscillation (~15 Hz) at .5 seconds in the cued condition. However, in the second-panel spectrogram at 0 seconds, there is a greater gamma band oscillation (~37-45 Hz). Therefore, in the last panel the gamma band oscillation was significantly larger for the no cue, and the beta band oscillation was significantly larger for the cued condition. Black boxes indicate the clusters of time-frequency bins that exhibited significant differences between conditions from the cluster-based permutation analysis, indicating that the difference did not occur by coincidence, at least in those black boxes. Both beta and gamma band oscillations are found in the “preparatory period” (time between the cue or no cue and the onset of speech).

For each cluster with a significant difference, paired t-tests were performed across all the cortical surface voxels to reveal source-level differences between conditions at each oscillation band (i.e., gamma and beta). The cortical surface map at the bottom of Figure 2 shows the distribution of t scores. The gamma band oscillation that was significantly larger for the no cue condition was located in the supratemporal gyrus of both the left and right hemispheres of the brain. The beta band oscillation that was significantly greater for the cued condition was located in the left hemisphere only of the inferior frontal gyrus.

## Discussion

Results from this study revealed that during a speech-in-noise task with a speaker identity cue, there is a greater activity of beta band oscillations in the inferior frontal gyrus of the left hemisphere of the brain (Figure 2). However, without a speaker identity cue for the target word presented in background noise, there is instead, greater activity of gamma band oscillations in the supratemporal gyrus in both hemispheres of the brain (Figure 2). These results were statistically significant and confirmed our hypotheses stating that the beta band oscillation would be present in the inferior frontal gyrus for the speaker identity cued condition. We also hypothesized that the gamma band oscillation will be located in the supratemporal gyrus for the condition with no speaker identity cue. While analyzing our EEG data in our study, we found physiological evidence that we use neurocognitive processing while listening in auditory complex environments. Beta band oscillations were dominated by top-down cognitive processes and gamma band oscillations were dominated by sensory sampling.

These findings align with past research that developed the concept of the Predictive Coding Theory. Our results support that when being surrounded by noisy environments and attempting to listen to target speech, the brain is using internal mental representations of the world and is generating predictions about where and what the sensory events are deriving from. If there is an error in the brain's prediction and there is a mismatch from its internal representations, then it is required to use sensory sampling (Arnal & Giraud, 2012; Friston 2005).  While interpreting our results and integrating them with past research findings, we know that if the listener is able to recognize the familiar voice and tracks the speaker's identity, then your brain is making predictions before receiving the sensory inputs of the words spoken by that same voice later on. Before the listener even hears the target sound, they prime their cortical representations to become biased towards the speaker's identity. Making predictions is a top-down process and applies prior knowledge from tracking the voice to fill in the gaps. In our study, we found beta band oscillations were present for the speaker identity cued condition (Figure 2). This aligns with work by Wang (Wang, 2010) that beta's cognitive function is to carry descending information and is dominated by top-down cognitive processing.

However, if there is not a familiar speaker, you are not able to make predictions or track the identity of the voice. This will cause prediction errors of the brain's hypothesis and there is a mismatch of the brain's internal representations of the world. Without the speaker identity cue, there was no tracking of the voice. Not being able to make predictions caused the brain to use sensory sampling to update the brain's hypothesis of what is occurring in its sensory environment. Findings from our current study contributed to the Predictive Coding Theory and the previous research findings because our results showed that gamma band oscillations were produced during the no-cue condition (Figure 2). Sensory sampling was occurring in this condition because the brain had to be open for multiple voices in the background noise. Sensory sampling is a bottom-up cognitive function and carries ascending information where the brain can interpret sensory signals found in gamma oscillations (Busch, Dubois, & VanRullen, 2009; Wang, 2010).

Findings from our current study found that gamma oscillation differences arose from the supratemporal gyrus in both hemispheres. The supratemporal gyrus contains sensory auditory cortices and is associated with the auditory sensory area. The cortical surface maps of the induced activities (Figure 2) during the no-cue condition further support that gamma band oscillations are elicited during sensory sampling. Sensory sampling occurs when there is no speaker identity to track because the listener utilizes frequent sampling of auditory features of all voices. In contrast, our results from the cortical surface maps (Figure 2) support that beta band oscillation differences arose from the left inferior frontal gyrus. The inferior frontal gyrus is also known as Broca's area, and is where higher levels of speech and language processing occur. Additionally, the left hemisphere specializes in speech and language functions. The inferior frontal gyrus is located very close to the motor strip of the cortex. Our research findings that showed beta oscillations in the area of the inferior frontal gyrus align with past research that supports that beta band oscillations specialize in motor functions. Motor and top-down processing are both used in speech and language tasks. You need prior knowledge to fill in missing information during communication barriers. Similarly, beta band oscillations are dominated by the cognitive function of top-down processing.

The timing of the induced activities of the cortical oscillations provides significant information. During this current study, we focused on analyzing the EEG data during the "preparatory period". The preparatory period was the timing interval between the presentation of the cue or no cue in the carrier phrase and when the onset of speech occurred with the target word in background noise. We found that gamma and beta oscillations were the most significant and had the greatest effect size during the preparatory period. This aligns with the work of Holmes et al. (Holmes, Kitterick, & Summerfield, 2018) because their findings showed that the increased duration between the cue-target interval in a multi-talker listening task allows for underlying preparation to take place over time. Our results supported this idea because during the preparatory period, the listener was able to prepare and prime their attention towards the target speaker identity. They were holding onto the valuable information of the male or female voice. However, with gamma, the listener would be preparing to sample multiple auditory features of multiple voices. Our results showed that top-down processing from beta waves and sensory sampling indicated through gamma waves occur after being cued or no-cued, but before listening to the target speech. This requires preparation and priming for future sensory events.

Through our current study, we have found what the “normative” brain does during a speech-in-noise task when they are provided with strong clues about what speaker identity voice they want to track. These findings can provide guidelines for what brain activities should look like in hearing-impaired listeners. In the future, we would like to use this information to make clinical advancements to help mitigate their communication barriers in noisy environments. These clinical implications will address their most frequent complaint and reduce their levels of strain and fatigue.

Limitations of our study include the small sample size. Additionally, there was a lack of behavioral evidence of the cueing effect. Individuals did not perform more accurately during the speaker identity-cue. Behavioral data showed low percentage scores of accurately choosing the correct target word during the task. This could perhaps be due to our choice of SNR of 0 dB, the task may have been too difficult for the listener. Our future works include testing clinical populations and investigating the neural substrates of individuals with hearing loss with this speaker identity research paradigm.

## Acknowledgements

This work was supported by the Department of Defense Hearing Restoration Research Program Grant (W81XWH-19-1-0637) awarded to Inyong Choi, as well as NIDCD P50 (DC000242 31). Further funding was generously provided by the Iowa Center for Research by Undergraduates (ICRU).

Special thanks to Dr. Inyong Choi, PI of the Human Auditory and Neuroscience Lab, for sharing his expertise in this field. His assistance in creating the experimental design, mentoring while conducting EEG analyses, and providing feedback while completing the manuscript, were all greatly appreciated. Additionally, we thank lab members; Jae Hee Lee, lab manager, and Mallory Orr, undergraduate research assistant, for both of their assistance in participant recruitment and data collection. Lastly, we thank Hwan Shim, for his assistance with EEG analysis, conducting the cluster permutation analysis, and creating the spectrogram and cortical surface map figures.

# References

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in cognitive sciences*, *16*(7), 390-398.

Best, V., Ozmeral, E. J., Kopčo, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *Proceedings of the National Academy of Sciences*, *105*(35), 13174-13178.

Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *Journal of Neuroscience*, *29*(24), 7869-7876.

Donner, T. H., & Siegel, M. (2011). A framework for local cortical oscillation patterns. *Trends in cognitive sciences*, *15*(5), 191-199.

Drennan, W. R., & Rubinstein, J. T. (2008). Music perception in cochlear implant users and its relationship with psychophysical capabilities. *Journal of rehabilitation research and development*, *45*(5), 779.

Engel, A. K., & Fries, P. (2010). Beta-band oscillations—signalling the status quo?. *Current opinion in neurobiology*, *20*(2), 156-165.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience*, *2*(10), 704-716.

Finke, M., Büchner, A., Ruigendijk, E., Meyer, M., & Sandmann, P. (2016). On the relationship between auditory cognition and speech intelligibility in cochlear implant users: An ERP study. *Neuropsychologia*, *87*, 169-181.

Fries, P., Nikolić, D., & Singer, W. (2007). The gamma cycle. *Trends in neurosciences*, *30*(7), 309-316.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, *360*(1456), 815-836.

Geller, J., McMurray, B., Choi, I., & Holmes, A. (2020, September 4). Validation of the Iowa Test of Consonant Perception. https://doi.org/10.31234/osf.io/wxd93

Gruber, T., & Müller, M. M. (2005). Oscillatory brain activity dissociates between associative stimulus content in a repetition priming task in the human EEG. *Cerebral Cortex*, *15*(1), 109-116.

Hochmair-Desoyer, I., Schulz, E., Moser, L., & Schmidt, M. (1997). The HSM sentence test as a tool for evaluating the speech understanding in noise of cochlear implant users. *The American journal of otology*, *18*(6 Suppl), S83-S83.

Holmes, E., & Johnsrude, I. S. (2020). Speech-evoked brain activity is more robust to competing speech when it is spoken by someone familiar. *bioRxiv*.

Holmes, E., Kitterick, P. T., & Summerfield, A. Q. (2018). Cueing listeners to attend to a target talker progressively improves word report as the duration of the cue-target interval lengthens to 2,000 ms. *Attention, Perception, & Psychophysics*, *80*(6), 1520-1538.

Jenkinson, N., & Brown, P. (2011). New insights into the relationship between dopamine, beta oscillations and motor function. *Trends in neurosciences*, *34*(12), 611-618.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063, pp. 218-226). Englewood Cliffs, NJ: Prentic Hall.

Koch, I., Lawo, V., Fels, J., & Vorländer, M. (2011). Switching in the cocktail party: Exploring intentional control of auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 11401147. https://doi.org/10.1037/a0022189

Lochmann, T., & Deneve, S. (2011). Neural processing as causal inference. *Current opinion in neurobiology*, *21*(5), 774-781.

Lu, Z. L., Tse, H. C. H., Dosher, B. A., Lesmes, L. A., Posner, C., & Chu, W. (2009). Intra-and cross-modal cuing of spatial attention: Time courses and mechanisms. *Vision research*, *49*(10), 1081-1096.

Mehraei, G., Shinn-Cunningham, B., & Dau, T. (2018). Influence of talker discontinuity on cortical dynamics of auditory spatial attention. *NeuroImage*, *179*, 548-556.

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., ... & Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research*, *351*, 68-79.

Richards, V. M., & Neff, D. L. (2004). Cuing effects for informational masking. *The Journal of the Acoustical Society of America*, *115*(1), 289-300.

Strait, D. L., & Kraus, N. (2011). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Frontiers in psychology*, *2*, 113.

Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *Journal of Neuroscience*, *16*(13), 4240-4249.

Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *Journal of Neuroscience*, *31*(25), 9118-9123.

Voisin, J., Bidet-Caulet, A., Bertrand, O., and Fonlupt, P. (2006). Listening in silence activates auditory areas: a functional magnetic resonance imaging study. *J. Neurosci.* 26, 273–278.

Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological reviews*, *90*(3), 1195-1268.

Wickham, M., Geller, J., & Choi, I. (2020). Effects of Speaker-Identity Cueing on Listening Effort During Speech-in-Noise.

Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: a remarkable past and a brilliant future. *Hearing research*, *242*(1-2), 3-21.

Zeng, F. G., Popper, A. N., & Fay, R. R. (Eds.). (2011). *Auditory prostheses: New horizons* (Vol. 39). Springer Science & Business Media.