# EEG differences between perceiving speech versus noise in physically identical sine-wave speech stimuli

A Thesis

Presented to

The Division of Philosophy, Religion, Psychology, and Linguistics

Reed College

---

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Arts

---

Camille Hendry

May 2019

Approved for the Division

(Psychology)

---

Michael Pitts

# Acknowledgments

I want to thank all the people who helped me make this possible. To my parents who have always encouraged me to try my hardest and never give up. To my sisters, Colette and Chloe for the constant support. My friend Amy Rose, Reed would not be the same without you. The SCALP lab for being my second home on campus. Finally, to Michael and Enriqueta both of who have shaped me into the scientist I am today.

# Table of Contents

# List of Tables

# List of Figures

# Abstract

Sine-wave speech (SWS) is a form of artificially degraded speech which has the unique quality of listeners initially perceiving it as noise, but after brief exposure to an undegraded version, the exact same SWS is readily perceived as speech. This makes SWS a great stimulus for studying the neural differences between speech and noise perception, because the physical input doesn't change, while perception changes dramatically. The perceptual switch from hearing noise to speech can also help test certain aspects of the predictive coding theory of the brain. Predictive coding is based on the notion that the brain might process the difference between a sensory input and a prediction based on prior knowledge and represent any mismatches as prediction errors which can then be minimized through an iterative process of updating the priors and retesting for mismatches.

There have only been five previous studies that have used SWS along with concurrent brain measures. However, in all of these studies, the SWS stimuli were task relevant which means that the neural findings could have been confounded by task effects. We designed an experiment to isolate the difference between speech and noise perception from task effects. The experiment consisted of three different phases, with the same exact physical stimuli presented in each phase while EEG data was recorded. In phase 1 the SWS was task irrelevant and perceived as noise. In phase 2, the SWS was also task irrelevant but was perceived as speech (due to a brief training between phases 1 and 2). In phase 3, the SWS was task relevant and perceived as speech. Data from a total of 18 participants was used for the main analyses and data from an additional 12 participants who spontaneously perceived the speech content of the SWS in phase 1 were used for additional control analyses. When comparing event related potentials (ERPs) elicited by the SWS in phase 2 vs. phase 1 a negative-going difference was observed over left fronto-central regions and was labeled the Speech Awareness Negativity (SAN). The SAN was not present for frequency flipped control stimuli that were always perceived as noise. When comparing ERPs elicited by SWS in phase 3 vs. phase 2 additional neural differences were observed, including a P3 component and a sustained frontal negativity

which can be attributed to tasks effects. Time-frequency analyses of the same EEG data were also conducted, and a suppression of alpha-band power was found in the SWS phase 2 vs. phase 1 comparison frontal regions and an enhancement of alpha-band power in the phase 3 vs. phase 2 comparison posterior regions. Overall, the results were consistent with the predictive coding framework, and the neural differences observed with this novel 3-phase paradigm serve as a useful starting point for refining our understanding of the neural mechanisms involved in basic aspects of speech perception.

# Introduction

All sensory information is ambiguous. It is the job of our sensory system to take this information and form some perception of the world that can help us interact with it. In the 1850s, Hermann von Helmholtz developed a theory known as unconscious inferences. This theory stated that human perception is being inferred from fragmentary data which require inferences from knowledge of the world to make sense of the ambiguous sensory signals (Gregory, 1997). Therefore, our sensory systems do not rely only on incoming information. Instead our perceptions of the world are more like well-informed hypotheses based on a comparison of current sensory inputs to previously stored information. Evidence for this theory includes situations in which our sensory systems interpret information incorrectly like in the case of certain perceptual illusions. In the case of illusions, just like in every day experiences, we are given ambiguous sensory information which we then form a hypothesis of based on our prior assumptions. However, various illusions are designed to provoke assumptions that directly contradict aspects of physical reality, i.e. our hypothesis is wrong, and we perceive something which is not inline with the physical input (see Figure 1).

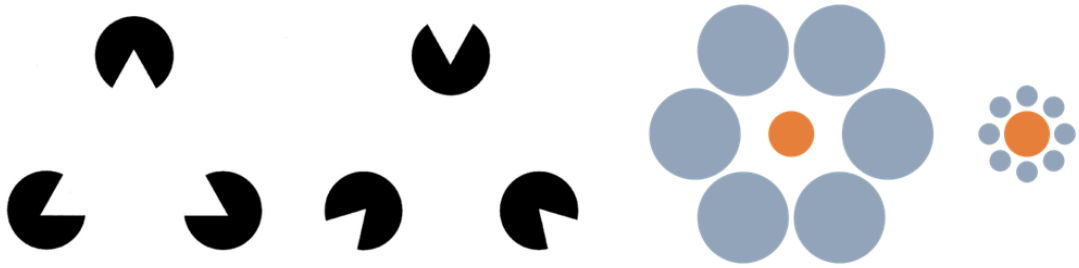

Figure 1: Optical Illusions (Brock, Brown, Boucher, & Rippon, 2002), (Gonzalez, n.d.) On the left is the Kaniza triangle. When the partial circles ("pacman" shapes) are turned inward, we perceive a triangle. Our visual system thinks it is more likely for a white triangle to be on top of three black circles than for three black circles with those exact pieces taken out to be arranged in such a way on a white background. On the right is a Gestalt proximity illusion (aka the "Ebbinghaus illusion"; aka "Titchener circles"). The orange circles are the same physical size. However, when you surround the orange circles with larger circles they typically appear smaller compared to when you surround them by smaller circles.

Helmholtz's theory relates to the idea of Bayesian statistics. Bayesian statistics is a mathematical procedure that applies probabilities to statistical problems (Friston, 2012). It has a fundamental view that a belief can change as new information is gathered, rather than being fixed based upon frequency or propensity (Aitchison & Lengyel, 2017). The idea that our brain implements some form of Bayesian statistics is known as the "Bayesian brain" hypothesis (Harkness & Keshava, 2017), meaning that neural populations represent sensory information probabilistically, in the form of probability distributions. This expands upon Helmholtz's original idea of our perception being more akin to hypotheses, by asserting that sensory processing is made up of conditional probabilities. Whichever probability has the highest likelihood determines our moment-to-moment perception of the world.

The Bayesian brain "attempts to explain why cognition produces the patterns of behavior that [it] does" (Jones & Love, 2011). However, it does not explain

mechanistically how the brain is able to achieve this. The theory of predictive coding provides a potential answer to this question of mechanism. Predictive coding is based on the idea that instead of representing the input directly, it is often more efficient to represent the prediction error, which is the difference between a sensory input and a prediction based on prior knowledge (Aitchison & Lengyel, 2017). Feed-forward connections convey stimulus-related information while feed-back connections provide the prediction. An example of a feed-forward pathway is the primary afferent pathway in the visual system (retina-LGN-V1-V2-V3 etc.)[1]. An example of a feed-back pathway in the same system starts with memory circuits and ends in early visual cortical areas (e.g., PFC/MTL-PTC-V3/V2/V1)[2]. It is the combination of these feedforward and feedback processes which forms the prediction error and, according to this theory, our perception results from the recursive minimization of this error signal. If a large enough error signal is elicited at any level of the system, the predictions from the next level are adjusted and iteratively tested through a feedback-feedforward cycle (see figure 2). The idea behind this theory is that it would be more efficient for a system with a rich history to code information in this manner (as opposed to freshly computing and representing all of the details of the incoming information at each moment in time) because the predictions will be correct most of the time, given a relatively stable physical world.

[1] LGN- lateral geniculate nucleus, V1-primary visual cortex, V2- secondary visual cortex, V3- third visual complex

[2] PFC- Prefrontal Cortex, MTL- medial temporal lobe, PTC- posterior temporal cortex

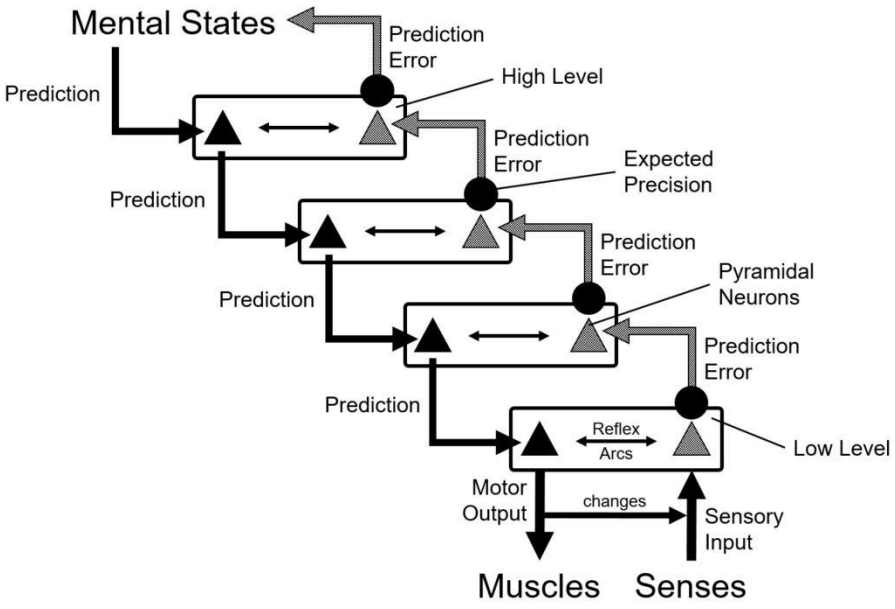


Figure 2: Predictive Coding Diagram ("The Bayesian Brain," 2018)
A diagram illustrating the hierarchal nature of a predictive coding system. At each level the prediction and the sensory information are evaluated to form a prediction error. It contains the feedforward process of sensory input, as well as the feedback process of prediction testing.

Although predictive coding hinges on the idea of prediction errors it does not specify how predictions are computed or how prediction errors are used. This is where the idea of the Bayesian brain can be combined with predictive coding. The predictions about the sensory input required by predictive coding may be based on conditional probabilities generated by Bayesian statistics.

Most of the research on predictive coding and Bayesian models of perception has focused on the visual system. This is because we know the most about this system and it is our most dominant sensory modality. However, it is important to keep in mind that these theories are designed to explain all types of perception, not just visual perception. The next most studied system is the auditory system. There is accumulating evidence for predictive coding in the auditory system. In short, over the past decade a broad range of findings in auditory neuroscience have pointed to a fundamental role of expectations and

prediction errors in auditory perceptual processing (Banai & Amitay, 2012). Although more research is necessary to test all theoretical elements of predictive coding theory, there is growing empirical evidence for this theory in the auditory domain.

# Brain Recording Methods

In order to understand studies involving the brain, it is important to understand the different methods for recording brain activity. Brief introductions to some of the leading brain recording methods, along with their strengths and limitations, are provided below.

## fMRI

Rather than looking at the structure of the brain via magnetic resonance imaging (MRI), functional MRI (fMRI) measures the amount of oxygen in the blood in different parts of the brain otherwise known as the Blood Oxygenation Level Dependent (BOLD) signal (Glover, 2011). When a population of neurons in a particular part of the brain is active (i.e. firing more action potentials), more oxygen is metabolized, leading to an increase in deoxygenated blood in that local area, and fMRI can detect this increase. fMRI has good spatial resolution (~1mm) and there is no need to inject radioactive substances as is required for positron emission tomography (PET) scans. fMRI can measure BOLD signals in individual "voxels" (volumetric pixels), and while the spatial resolution is high, each voxel typically contains over 500,000 neurons. fMRI has low temporal resolution due to the time it takes for blood oxygen levels to change due to increased neural firing (~2-6 sec), while neural processing itself is known to occur on a much faster time-scale (10th of msec). Thus, fMRI is not a good choice for recording brain activity if you want to know the time point at which a certain neural event is happening. It is also one of the most expensive techniques in terms of equipment and operation costs (~3 million USD).

## ECoG

Electrocorticography (ECoG) is a type of electrophysiological monitoring that uses electrodes placed directly on the exposed surface of the brain (or deep within brain tissue) to record electrical activity from the cerebral cortex (Hill et al., 2012). This electrical activity includes post-synaptic dendritic currents (the immediate electrical after-effect of neural transmission across a synapse), local field potentials, and extracellular "spikes" from single-neuron action potentials. ECoG requires a craniotomy (a surgical incision into the skull) to implant the electrode grids. Since ECoG is such an invasive procedure it used exclusively in patients who are already having surgery for other reasons (e.g., severe epilepsy). This can lead to a limited number of participants in an ECoG study. Also, electrodes tend to be placed in different locations from person to person, so this spatial variability in data acquisition has to be accounted for in any group-level analyses.

## MEG

Magnetoencephalography (MEG) is a non-invasive functional neuroimaging technique for mapping brain activity by recording magnetic fields produced by electrical currents occurring naturally in the brain (post-synaptic dendritic currents), using very sensitive magnetometers (Singh, 2014). MEG provides good temporal resolution (1 msec) as well as decent spatial resolution (~1-10mm), and excellent spatial coverage (whole brain). However, MEG can only record activity from neural populations that are oriented parallel to the cortical surface, and only from superficial sources that are close to the sensors, leaving lots of neural activity unmeasured (perpendicularly aligned neural populations and deep sources). Similar to fMRI, MEG has a high cost of equipment and operation (~$3M).

# EEG/ERPs

Electroencephalography (EEG) measures the post-synaptic electrical potentials generated by large populations of synchronously active neurons (Tivadar & Murray, 2019). This is done by placing electrodes on the scalp, amplifying the signal and then plotting changes in voltage over time, relative to a reference electrode. Unlike ECoG, EEG is noninvasive and therefore can be used on a larger variety of subject populations, including healthy volunteers, infants, and individuals with various neurological disorders. The clear strengths of EEG are that it provides excellent temporal resolution (~1 msec), full spatial coverage (whole brain), and can measure signals from both deep and superficial sources in populations of neurons oriented in any axis relative to the recording electrode. However, it does not provide great spatial resolution (~1 cm), due to the smearing of the electrical signals by the skull, scalp, meninges, CSF, etc., and similar to MEG, cannot measure activity in certain areas of the brain (because of non-parallel arrangements of neurons that prevent the summation of post-synaptic currents). This technique is useful for studies in which the focus is on the stages of processing more than which specific bran areas are involved. EEG also has the lowest cost in equipment and operating cost (~$60k) making it more accessible and easier to acquire data from a larger number of subjects.

When analyzing EEG data it is important to understand that the raw data is not very useful for most purposes. The raw electrical signal is a combination (superposition) of multiple different brain processes that are happening simultaneously. However, within the EEG exist event related potentials (ERPs). ERPs are neural responses that are related to specific (cognitive, motor or sensory) events. When experimental designs incorporate repeated trials of the same events, one can average across the multiple occurrences of the event to isolate the ERP from the rest of the EEG. By conducting various controlled experiments, researchers have discovered many different ERP components that are each associated with specific sensory, cognitive, or motor processes (Luck & Kappenman, 2011). These ERP components can then be used as “tools” in subsequent experiments to answer broader questions in experimental psychology (Luck, 2014).

Another way of analyzing ERP data is known as time-frequency analysis. The raw electrical signal we see is made up of a combination of different frequencies into one waveform (Figure 3). Time frequency analysis allows you to assess what frequencies were present or absent, or stronger or weaker, across time. One of the disadvantages of ERPs is that they only pick up brain activity that is consistently phase-locked to the onset of the stimulus. Time-frequency analysis can reveal event-related oscillations that are not phase-locked because the wavelet analysis is done on individual trials rather than on trial-averaged data.

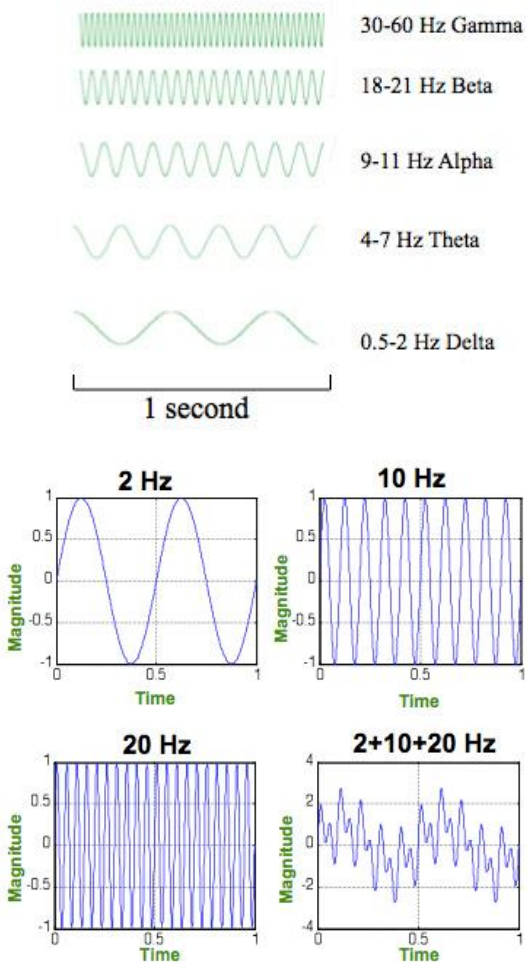


Figure 3: EEG Frequencies ("Time frequency tutorial - SCCN," n.d.)

The top graph shows the diffeent frequencies types of EEG data across one second. The bottom four graphs depict how different frequencies combine to form an EEG waveform. The final waveform has a combination of delta, alpha and beta frequencies within it.

# Speech Perception

Speech perception is an excellent topic for investigating how our brains might use predictive coding to process sensory information. Speech is a complex natural stimulus that is often ambiguous due to substantial acoustic variation across speakers. Speech perception relies heavily on contextual factors such as acoustics, syntax, and semantics, which can all contribute to priors that are compared with the incoming auditory information. This section provides a brief overview of the basics of speech stimuli, the brain areas known to be involved in processing such stimuli, and a consideration of the importance of studying speech perception within the predictive coding framework.

Speaking involves the production of meaningful streams of sounds. A spectrogram shows the patterns of frequency and amplitude that ground audible features (Cassidy, 2002)(see Figure 4).

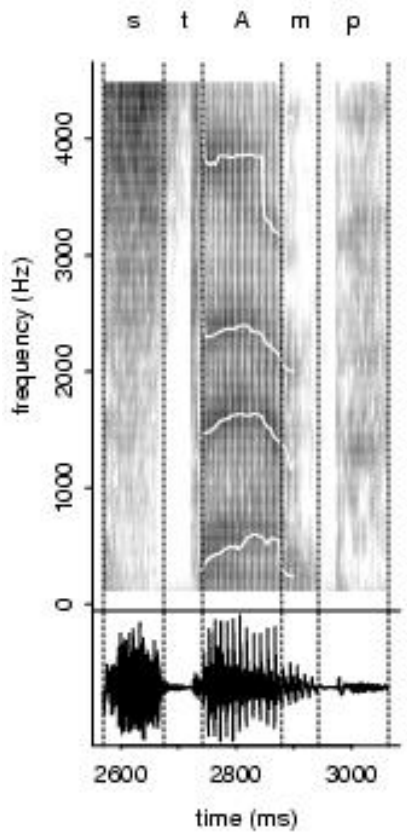


Figure 4: Speech Spectrogram (Cassidy, 2002)

The speech spectrogram of the word “stamp” with the sound waveform on the bottom.

While a spectrogram of speech shows complex acoustic structure involving patterns of audible qualities over time, the physical structure of speech depicted in a spectrogram is often quite different than the structure of perceived speech. For example, speech is perceived to be segmented in different ways than the acoustical structure of a spectrogram would predict. The most salient perceptual segments are words, but speech can also be broken up into smaller perceptual segments such as phonemes, the smallest unit of speech that changes the meaning in a word. In English there are 47 phonemes, 13 major vowel sounds and 24 major consonant sounds (Goldstein, 2009). Vowel sounds are caused by a resonant frequency of the vocal cords and produce peaks in pressure at a number of frequencies called formants. The first formant has the lowest frequency, the second has the next highest (Samuel, 2010), etc. Formants occur at roughly 1000Hz intervals and each corresponds to a resonance in the vocal tract. Although all vowel sounds have at least four formants, the two first formants are usually all that are needed for one to disambiguate vowel sounds. Four formants can be seen in Figure 4.

Two areas of the brain which are known to be important for speech processing are Broca's area and Wernicke's area (Samuel, 2010). Broca's area, located in the inferior frontal lobe (most often in the left hemisphere), is often considered to be the major speech production area of the brain, and it communicates directly with motor areas that control movements required to produce speech sounds. When there is a lesion to this area, patients can understand the speech of others, but have difficulty producing meaningful speech themselves, i.e. "Broca's aphasia" (Fridriksson, Fillmore, Guo, & Rorden, 2015). Wernicke's area, located in the superior posterior temporal lobe near the intersection with the inferior parietal lobe, is considered to be the major speech comprehension area of the brain, and receives direct input from auditory sensory areas within the temporal lobe. If Wernicke's area is lesioned, patients develop "Wernicke's aphasia". This is characterized by the inability to understand spoken or written word. Unlike Broca's aphasia, Wernicke patients speak fluently, but the speech that is produced is often meaningless gibberish (Hartman et al., 2017).

Speech perception is a prime example of our sensory system having to disambiguate sensory information. We are able to pick out speech from other sensory

inputs when we are in a noisy environment. If speech perception follows Bayesian principles, then it must involve comparing current auditory input to our prior knowledge of speech and making a probabilistic prediction. That prediction then shapes our perception. At the most basic level, we are even able to identify speech stimuli as speech when it is a language we are not familiar with. Although we do not have knowledge to interpret the speech, we are able to make a prediction that it is speech based on our prior knowledge of the physical characteristics of speech stimuli that we have learned to understand. Several previous studies have provided clear evidence of predictive coding during speech perception. For example, in one study participants were presented degraded speech stimuli and their ability to recognize the stimuli was tested. Participants were then trained to understand the speech content in the degraded sounds by viewing written versions of the words (giving prior knowledge). After this training they had a higher rate of speech recognition (Sohoglu & Davis, 2016). This evidence is consistent with the predictive coding framework because the written words provided information to update the priors which then helped to minimize the prediction error. This study also used EEG and MEG to gather temporal and spatial information about the underlying brain activity elicited by the speech stimuli both before and after training. The main finding was that when speech clarity was enhanced during the training phase through prior knowledge, they observed reduced neural activity in a peri-auditory region of the superior temporal gyrus (STG). The perceptual learning effect (pre vs. post training) also reduced activity in a nearly identical region of the STG leading them to conclude that they were working on the same mechanism.

## Sine-wave Speech

Sine-wave speech (SWS) is a form of artificially degraded speech first developed at the Haskins Laboratory [3] (Remez, Rubin, Pisoni, & Carrell, 1981). Before the listener is aware of the speech content within SWS stimuli they will typically perceive the SWS as noise. However, once listeners are informed of the speech content, the exact same

[3] Link to example SWS: https://www.mrc-cbu.cam.ac.uk/people/matt.davis/sine-wave-speech/

SWS stimuli are readily perceived as speech. The physical stimuli themselves remain the same before and after the brief training, while one's perception radically changes from "noise" to "speech". These features render SWS ideal stimuli for investigating neural differences between speech and noise perception, as perceptual differences can be cleanly isolated from physical (acoustical) attributes of the stimuli.

SWS is generated by using a formant tracker to detect the formant frequencies found in an utterance, and then synthesizing sine waves that track the center of these formants. The formants are then replaced by pure tone whistles. (See Figure 5)

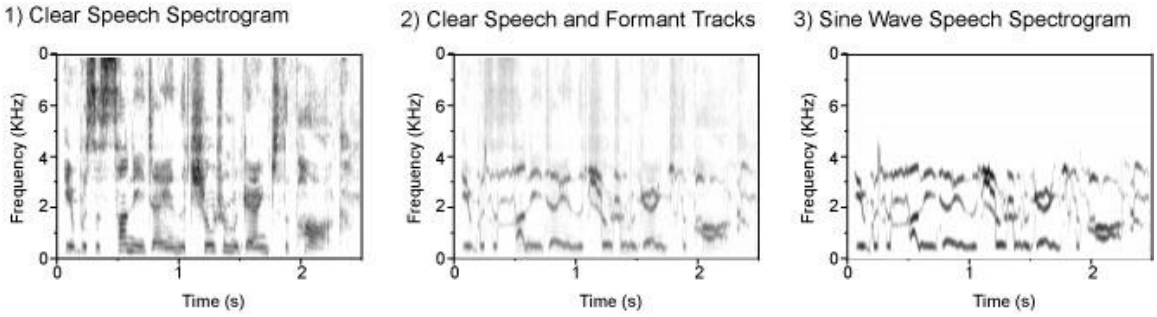


Figure 5: Sine-Wave Speech Generation (Davis, n.d.)
Image 1 depicts the raw speech spectrogram. Image 2 shows the formants of the speech being tracked. Image 3 shows the synthesis of these formants to form sine-wave speech.

Most previous experimental paradigms for studying the neural basis of speech perception have compared brain activity elicited by speech stimuli to brain activity elicited by distorted speech, speech of another language, pseudowords, and other physically altered non-speech stimuli (Samuel, 2010). However, in all of these studies, it was unclear whether the measured neural differences were due to speech vs. non-speech *perception* or to the different acoustical inputs to the auditory system, i.e. there was always an underlying confound of comparing brain activity elicited by two physically different stimuli. Experimental paradigms that use SWS avoid this critical confound by creating a comparison of speech to non-speech perception with a physically unchanging stimulus. The only changes are in the listeners' priors and their resulting perception of the stimulus.

# Previous SWS Studies

Although the nature of SWS makes it an ideal stimulus to study neural mechanisms of basic speech perception, very little cognitive neuroscience research has utilized SWS. Only five papers have been published so far using SWS as a stimulus along with a brain recording technique (Dehaene-Lambertz et al., 2005; Khoshkhoo, Leonard, Mesgarani, & Chang, 2018; Liebenthal, Binder, Piorkowski, & Remez, 2003; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Möttönen et al., 2006). In the following, I provide brief summaries of this previous work.

The first study that investigated the neural mechanisms of speech perception using SWS was Liebenthal et al. (2003). Participants were given an auditory task during an fMRI scan. The auditory task consisted of determining whether an isolated second-tone formant was included in a three-tones sinewave complex. The second-tone formant was either aligned making a SWS word or temporally reversed relative to the other tones in the complex control stimulus. Participants were unaware of the phonetic aspect of the SWS in the first half of the experiment. Before the second half of the experiment participants were trained to understand the SWS. After training they saw an increase in behavioral response time before returning to times similar to pre-training and a decrease in left Heschl's gyrus activation.

Liebenthal et al. (2005) set out to further investigate the neural substrates of phonemic perception. To do this they compared fMRI activity elicited by English syllables to SWS speech versions of the syllables. Throughout the experiment the majority of the participants only heard the SWS as nonfamiliar syllables. Participants had to perform a two alternative forced choice ABX discrimination task where they determined which syllable (A or B) was most closely related to a test syllable (X). Participants performed worst behaviorally on the SWS or nonfamiliar syllable task. They also found an area extending along the left middle and anterior superior temporal sulcus that was more responsive to familiar consonant–vowel syllables than to the SWS syllables which could not be associated with learned phonemic categories. One of the main limitations of this study was that they did not utilize the perceptual switch which can occur with SWS stimuli. Instead they compared SWS to normal speech. While this

may be a more controlled comparison than between words and non-words there is still a physical difference between the stimuli.

The next SWS study, conducted by Dehaene-Lambertz et. al. (2005) investigated the neural difference when participants switched from noise to speech perception. They utilized both ERPs and fMRI in a discrimination paradigm. Their paradigm consisted of three phases. The first phase was a passive phase where participants watched a silent movie and were asked to ignore the stimuli. The results of this phase were not reported. In the second phase participants were presented SWS syllables but perceived them as noise. They were tasked to press a button when the stimulus had changed. Since they were not perceiving speech, they were essentially performing an auditory discrimination task on different "noises". Between the second and third phase participants were briefly trained to hear the SWS as syllables. In the third phase they performed the same task, but now since they could hear the stimuli as speech, they were performing a phonetic discrimination task. The results of the study showed that the electrophysiological mismatch response occurred earlier for a phonemic change than for a physically equivalent acoustic change. This result led the authors to conclude that phonetic coding is faster than acoustic coding. In fMRI, they found that hearing the SWS stimuli as speech enhanced activation in the posterior parts of the left superior temporal gyrus and sulcus compared to perceiving the same stimuli as noise. They also found that activity within the thalami, basal ganglia, insula and frontal operculum in mainly the left hemisphere was enhanced in trials where participants responded "different" on the task. This study was able to provide evidence that although the stimulus never changed, the change in perception of SWS was processed differently in the brain.

A study by Möttönen et al. (2006) attempted to address some of the limitations of the Liebenthal et al. (2005) study. In this two-session fMRI study, participants were tasked with identifying which of three different sounds were presented. Two of the three sounds were SWS pseudowords and one was a control. During the first session participants perceived all three sounds as non-words. Between the first and second session participants were taught how to perceive the SWS pseudowords as speech. During the second scan, the stimuli and the task were identical to the first scan, while

participants' perceptions of the SWS now differed. The results showed that the SWS stimuli elicited significantly stronger activity within the left posterior superior temporal sulcus in the second session, when they were perceiving speech, compared to the first session, when they were perceiving the same stimuli as non-speech. Importantly, the control stimuli had similar activity in this region in both sessions, ruling out a potential explanation of the neural changes due to exposure, learning, or condition order. These findings support the results of the Liebenthal et al. (2005) and the Dehaene-Lambertz et al. (2005) studies, suggesting that the left posterior superior temporal sulcus plays a key role in basic speech perception.

The most recent SWS study, by Khoshkhoo et al. (2018), used ECoG to investigate the cortical representations of SWS. Three patients were implanted with high-density multi-electrode cortical surface arrays. In the first phase of the experiment, participants passively listened to 24 different SWS sentences. Each sentence contained one color word and one number word. One of the three participants was able to recognize the speech content of the SWS during the first phase, while the other two participants perceived the stimuli as noise. In the second phase of the experiment participants listened to the original speech versions of the 24 sentences and were tasked with identifying the color and number words in each sentence. Prior to phase three, they were informed that the SWS sentences were modified versions of the sentences from phase 2. They then performed the same identification task as phase 2. The results of this study showed that when the SWS was not perceived as speech only the auditory cortex discriminated speech sounds. However, when SWS was understood as speech the inferior frontal cortex also discriminated speech sounds, in addition to auditory areas within the temporal lobe.

Across these five studies most of them implemented either a discrimination task, choosing whether it was a SWS stimulus or a control stimulus, or an identification task, in which listeners had to identify a specific portion (syllable or word) within the SWS stimuli. The SWS stimuli ranged from a syllable, to a word or pseudoword, to full SWS sentences. Several studies reported more activation in the left superior temporal sulcus and gyrus when the SWS stimuli was being perceived as speech compared to noise. One study also reported activation in the inferior frontal cortex for this same comparison.

# The Current Study

All of the previous SWS studies used paradigms in which neural comparisons were made between task irrelevant SWS when perceived as noise vs. task relevant SWS when perceived as speech, or task relevant SWS when perceived as noise vs. speech. This means that the previous experiments could be confounded with brain activity associated with task relevancy, as one of the key conditions in each comparison always included task-relevant SWS. In other words, it could be that some of the neural differences observed were due to neural processes related to completing the task rather than perceiving the stimuli. A similar issue has been raised for studies trying to isolate the neural correlates of consciousness (NCC) (Aru, Bachmann, Singer, & Melloni, 2012). Similar to studies that compare brain activity for perceived vs. not-perceived stimuli, previous SWS experiments may have captured both perceptual and post-perceptual processes in the neural contrasts (with no way of differentiating between the two), rather than isolating processes directly linked with speech perception.

In order to address these criticisms, we conducted a SWS experiment designed to isolate neural differences linked with speech perception from those related to the task. The experiment consisted of three different phases, with the same exact physical stimuli presented in each phase while EEG data was recorded. In all three phases, three different SWS speech words were presented along with frequency flipped control versions of each SWS word, as well as pure tones of three different frequencies. Participants were initially told that the SWS and control stimuli were randomly generated computer noise. In phase 1 participants performed a one-back task on the pure-tones of varying pitch (pressing a button whenever they heard the same pure-tone stimulus repeat twice in a row). After Phase 1 participants filled out a questionnaire to see whether they perceived any speech content in the SWS stimuli. A majority of subjects (18 out of 30) did not perceive any speech in phase 1. In between phases 1 and 2, participants were trained on the SWS stimuli in order to induce the perceptual shift in perceiving the SWS as words. Phase 2 was identical to phase 1 in stimuli and task (i.e. the SWS stimuli remained task-irrelevant), with the only difference being how the SWS was perceived (due to the

intervening training). Once again, after phase 2, subjects were given a post-phase questionnaire to see if they perceived words in the SWS. All participants (except one) reported hearing the SWS stimuli as speech in phase 2, and 17 out of 18 were able to correctly identify the three words presented. Phase 3 had the same stimuli as phase 1 and 2 but the task was changed such that participants now had to perform the one-back task on the three SWS words, thus rending these stimuli task relevant in this last phase.

By comparing brain activity elicited by task-irrelevant SWS stimuli in phase 1 (perceived as noise) vs. 2 (perceived as speech), we sought to better isolate neural correlates of speech perception from post-perceptual task-related processing. By including phase 3, in which the SWS stimuli were task-relevant, we were also able to compare our results to previous studies that may have confounded speech perception and task.

# Methods

## Participants

Thirty-three people (aged 18-23) with normal or corrected-to-normal vision and no history of brain injury participated in this study. Participants were compensated $20 for their time. All procedures were approved by the Reed College Institutional Review Board**.**

## Stimuli

A total of 9 auditory stimuli were used throughout the main experimental phases of this study: three SWS words (brain, wave, yard), three control versions of these words, and three pure-tones. Additional SWS words were used in the intervening training between phases 1 and 2. The words brain, chill, church, language, speech, wave, world, yard and zombie were spoken into a microphone by a male voice. To create the sine wave speech versions of these words we used Praat software, which utilized a formant tracker to detect the formant frequencies of each word, and then synthesized sine waves that track the center of these formants thereby creating sine wave speech (Boersma, 2002).

The control stimuli were created for the three SWS words used in the main phases of the study: brain, wave and yard. They were created by inverting ("flipping") the spectral frequencies of the first two formants of each word, while maintaining the other attributes of the soundwave (e.g. energy, movement, average frequency, etc.).

The three pure tone stimuli ranged from low pitch (~500 Hz), medium pitch (~1250Hz) and high pitch (~2000 Hz). The pure-tones were 600ms in duration, while the SWS stimuli ranged from ~480 to 600ms in duration. All stimuli were presented at ~46.5dB. To help subjects avoid eye movements, a fixation dot (1 deg visual angle) was presented constantly throughout the experimental trials.

# Equipment

The electrophysiological data was collected using a 96-channel passive electrode system (Acticap) (Appendix C). The data was acquired using EEG Recorder software and analyzed via Brain Vision Analyzer software (Brain Products, Herrsching, Germany). Stimulus presentation was controlled via Presentation software (Neurobehavioral Systems, San Francisco, USA). All stimuli were presented in stereo to the left and right ears using ER-2 research-quality headphones. These were equipped with flat frequency response at the human eardrum, 70+ dB isolation between ears and 30+ dB external noise exclusion.

# Procedure

All participants attended one ~3 hour recording session. The first thirty to forty-five minutes of the session were used to fit and prepare the electrode cap for EEG recording. Electrode gel was applied at the site of each electrode and a wooden stick was used to lightly abrade the scalp.

Participants then performed computerized tasks from the main experiment which was broken up into three different phases (see figure 6) (each phase lasted ~18 min; short rest breaks were provided every ~2-3 min, with longer rest breaks between each phase). The phases only differed in what participants perceived and what task they performed, while the stimuli and way in which the stimuli were presented remained the same across all phases. For each phase, the 9 stimuli (3 SWS, 3 control, 3 tones) were presented 100 times, for a total of 900 trials. Each trial consisted of a ~600ms stimulus, followed by a 500-700ms silent Inter Stimulus Interval (ISI) (total trial duration = ~1200ms). Throughout the entire experiment a fixation dot was present on the screen and participants were instructed to fixate on it for the duration of the experiment. 30% of trials for each of the 9 stimuli were “one-back” trials in which the same stimulus repeated back-to-back. These one-back trials were excluded from EEG analysis but allowed for

the main task manipulation (pure-tone one-backs were targets in phases 1 and 2; SWS one-backs were targets in phase 3).

Before starting the three phases participants were given a practice block. The practice block started with the presentation of each of the 9 stimuli to make sure the participants could hear them. The volume of the stimuli was kept the same across all participants. The practice block had the same task as phase 1: press a button for every pure-tone one-back trial. Participants were told that the other stimuli (SWS and control stimuli) were randomly generated computer noise. After the practice block, participants moved on to phase 1. Participants were given the awareness questionnaire (Appendix A) after the completion of phase 1. The awareness questionnaire served to identify whether they had noticed the speech content in the SWS stimuli. If participants did not report hearing speech on the initial open-ended question and marked a 3 or lower on the confidence scale for "hearing distorted words", then they were considered to not have noticed the speech in phase 1. If they reported hearing speech on the open-ended question or marked a four or five on the confidence scale for "hearing distorted words", they were considered spontaneous speech perceivers and asked to write down as many words as they remembered hearing. These groups completed the rest of the experiment in the same way but were treated differently during the analysis of the data.

After completing the questionnaire, participants completed a brief passive training session. The passive training exposed the participants to nine different SWS stimuli including the three target SWS stimuli. The SWS word was played, followed by the plain English version, and then the SWS word again. Each SWS-word-SWS sequence was repeated 3 times, and then participants assessed whether they could easily hear the sine wave word as speech. If they were not easily hearing it as speech, they could repeat the sequence as often as they needed, before moving on to the next word. The participant then completed a brief active training session. In the active training participants had to identify whether a given word was SWS or noise (control stimulus). If the stimulus was a SWS word, they had to choose which specific word it was from nine options presented on screen; if it was a control stimulus, they could press a tenth "no word" button. They were presented each of the nine SWS words (3 times each) along with the three SWS controls (which should have been identified as noise), all in random order.

After completing the training sessions, participants started phase 2. Phase 2 consisted of the same stimuli and task as phase 1. After phase 2 participants had to fill out the same awareness questionnaire again as well as a speech and recall questionnaire (Appendix B). First a free recall question in which participants had to list any words they heard in the prior phase. A second recognition question contained a list of words consisting of the three targets words, three words presented in the training, and three words never presented in the study. Participants were instructed to circle any words they heard in phase 2 of the study.

Finally, participants completed phase 3. This final phase had the same stimuli as phase 1 and 2, but with a change in task. Participants were now instructed to perform the one back task on the 3 SWS stimuli instead of the pure-tones. There was no questionnaire given after phase 3.

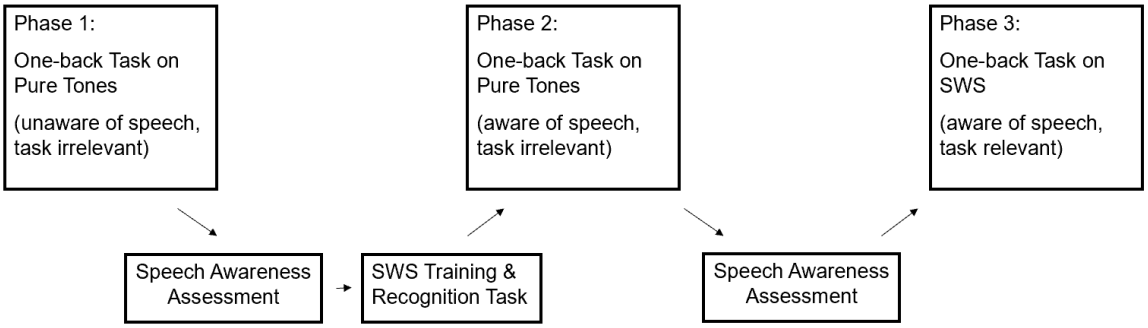


Figure 6: Methods Diagram

The general procedure of the experiment. There will be three main phases with either the perception or the task changing between phase. Between phase l and two there is a speech awareness questionnaire along with two training tasks. Between phase 2 and phase 3 the questionnaire is repeated.

# Data Analysis

Only trials which did not contain a response were included in analysis. The data was segmented into nine groups based on stimulus type and phase. They were filtered

with high-cut offs at 25Hzts. An average reference was formed using all electrodes. Artifacts (blinks, eye movements, facial muscle noise, etc.) were rejected semi-automatically (on average 22.84% of trials were rejected due to artifacts across all conditions). Two participants were excluded due to having more than 50% of trials rejected due to artifacts. All the individual segments of each group were averaged together. All segments each ERPs were time locked to main speech envelope for each stimulus (see appendix D) and baseline corrected.

Due to the novel design of this study there was no ERP component that could be targeted a priori to measure the perceptual shift that occurs when SWS stimuli are perceived as speech vs. noise. Based on an initial visual inspection of the data, we identified a clear negativity over left frontal electrodes at ~200-300ms when comparing the ERPs for SWS in phase 2 (perceived speech) versus phase 1 (perceived noise) in the grand average of subjects who did not spontaneously hear the speech in phase 1. We then ran two different control analyses at this same time window and electrodes, (1) ERPs elicited by control stimuli in phase 2 vs. 1 (these stimuli were perceived as noise in both cases), and (2) ERPs elicited by SWS stimuli in phase 2 vs. 1 in the group of subjects who spontaneously perceived speech in phase 1 (these stimuli were perceived as speech in both cases). We also analyzed the well-known ERP component called the P3 (also known as P300 or P3b) across all three phases.

In addition to the main ERP analyses, we also ran time-frequency analyses across the three phases for the SWS and control stimuli. All the data was filtered with a 60Hz notch filter. The data was then segmented by phase and stimuli. A continuous wavelet transform was performed using a Morlet complex. There was a frequency range of 1Hz-40Hz and a Morlet parameter of 6. Wavelet normalization was done through instantaneous amplitude. The baseline was corrected based on a -350ms to -50ms time window. Again, we did not have a priori predictions, so we used initial visual inspection of the grand average data to identify areas of interest, and then followed this up with several control analyses. When looking at the SWS event-related spectra in phase 2 versus phase 1 we focused on a left frontal region which showed spectral power differences in the high alpha low beta range (~10-15 Hz), from 100-200ms. For SWS

phase 3 versus phase 2 we focused on a right central region in the alpha range (8-13 Hz), from 150-250ms.

# Results

## Behavioral Data

Eighteen participants were unaware of the speech content within the SWS stimuli in phase 1, while twelve had some level of speech awareness in phase 1. These two groups of subjects will be referred to as "non-noticers" and "noticers" from this point on. The main focus of the EEG analysis will be on the non-noticers group. One participant was removed from further analysis for not hearing any words in phase 2.

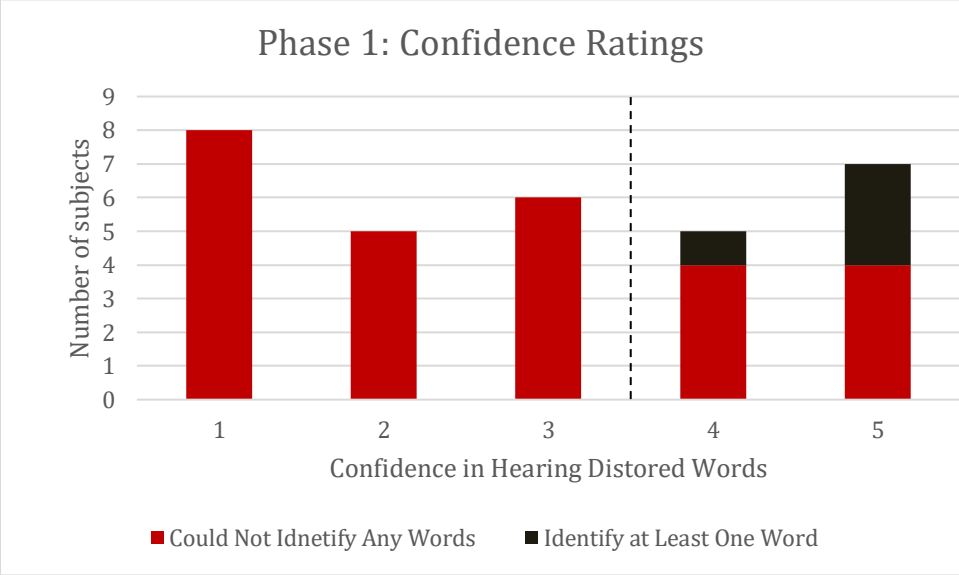


Figure 7: Phase 1 Confidence Ratings

The frequency that participants choose for their confidence in hearing distorted words in phase 1. Anyone who is on left of the dotted line were considered to not have perceived speech and form the "non-noticers" group. Everyone to the right of the dotted line were considered to hear speech and put in the "noticers" group.

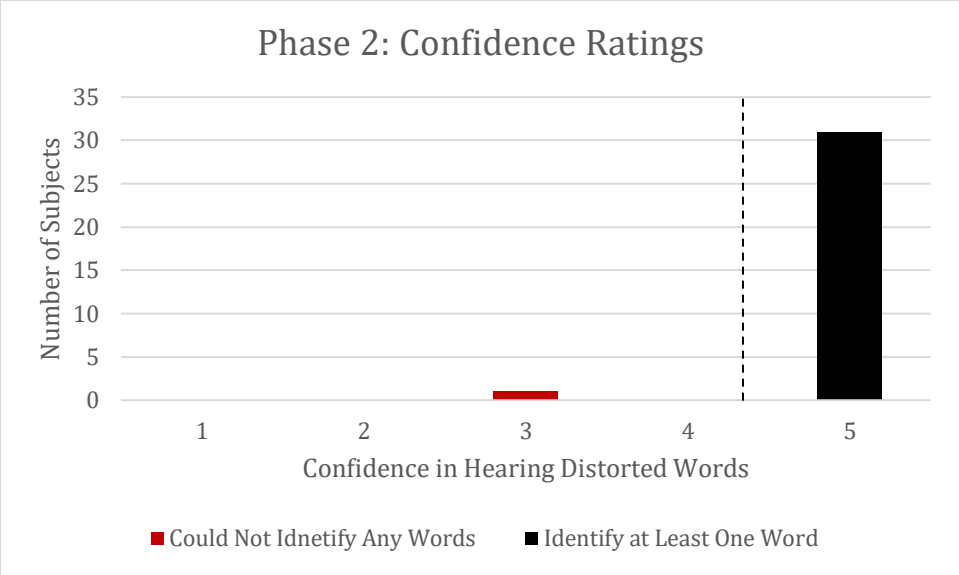


Figure 8: Phase 2 Confidence Ratings

The frequency that participants choose for their confidence in hearing distorted words in phase 2. Anyone who is on left of the dotted line were considered to not have perceived speech and anyone to the right of the dotted line were considered to hear speech. All subjects were meant to r speech during this phase so anyone to the left of the dotted line were exclude from the study.

| Group | Phase 1 | Phase 2 | Phase 3 |
|---|---|---|---|
| Non-Noticers | 97.42 | 96.5 | 88.1 |
| Noticers | 99.13 | 96.85 | 92.8 |

Table 1: One-back task accuracy of both groups across the three phases

| Group | SWS Target Words | Other SWS Words | Non-word |
|---|---|---|---|
| Noticers | 100% | 93.52% | 85.80% |
| Non-Noticers | 97.90% | 97.47% | 86.87% |

Table 2: Active training (Recognition Test) accuracy across both groups

# Electrophysical Results

## ERPs

### *Speech Awareness Negativity (SAN)*

For the non-noticer group of subjects, eight electrodes (5, 6, 16, 17, 18, 31, 32, 33) (see appendix C for electrode map) over the left frontocentral scalp were pooled and the mean amplitude was measured from 200-300ms. A 2 (stimuli: SWS, control) by 3 (phase: 1, 2, 3) repeated measures ANOVA with a Greenhouse-Geiser correction revealed a main effect for phase ($F=8.27$, $p= 0.0035$) and an interaction between phase and stimuli ($F=12.39$, $p=0.0002$), but no main effect of stimuli ($F=0.26$, $p= 0.6180$). A follow up pair-wise comparison revealed that the interaction was due to a significant difference in ERP amplitude between SWS phase 2 and SWS phase 1 ($t=3.09$, $p=0.004$), and SWS phase 3 and SWS phase 2 ($t=-6.78$, $p = 0.00$). Pair-wise comparisons for control stimuli in phase 2 vs. phase 1 ($t=-1.73$, $p = 0.093$) and phase 3 vs. phase 2 ($t=0.06$, $p = 0.951$) showed no amplitude differences.

A follow-up analysis was also conducted with the 12 noticer subjects. A 2 (stimuli: SWS, control) by 3 (phase: 1, 2, 3) repeated measures ANOVA with a Greenhouse-Geiser correction revealed a main effect for phase ($F=6.45$, $p=0.0212$), but no main effect of stimuli ($F=2.38$, $p=0.1508$) or interaction between phase and stimuli ($F=1.99$, $p= 0.1748$).

Because this ERP difference was only detected when perception changed from noise to speech (and was absent in the two control comparisons, as well as in the comparisons across tasks), and because it led to more negative-going amplitudes when speech was perceived, we theretofore refer to this ERP effect as the “speech awareness negativity” or “SAN”.

### *P3*

Eight electrodes (13, 14, 27, 28, 29, 44, 45, 46) were pooled and the mean amplitude was obtained at 400-600ms. A 2 (stimuli: SWS, control) by 3 (phase: 1,2,3) repeated measures ANOVA with a Greenhouse-Geiser correction revealed a main effect for phase (F=6.35, p=0.0053) and an interaction between phase and stimuli (F=4.47, p=0.0346). A follow up pair-wise comparison was run a found a significant difference between SWS stimuli phase 3 vs phase 2 (t=3.81, p=0.001) and SWS stimuli phase 3 and phase 1 (t= 3.66, p=0.001). There was no amplitude difference for control stimuli phase 3 vs phase 2 (t=-0.37, p=0.714) and control stimuli phase 3 vs phase 2 (t=0.99, p=0.329).

A follow-up analysis was also conducted with the 12 noticer subjects. A 2 (stimuli: SWS, control) by 3 (phase: 1, 2, 3) repeated measures ANOVA with a Greenhouse-Geiser correction revealed no main effects for stimuli (F=4.40, p= 0.0599), a main effect of phase (F=9.63, p= 0.0011) and no interaction between stimulus and phase (F=2.74, p=0.1111).

### *Frontal Negativity*

Eight electrodes (2, 3, 7, 8, 9, 20, 21, 22) were pooled and the mean amplitude was obtained at 400-600ms. A 2 (stimuli: SWS, control) by 3 (phase: 1, 2, 3) repeated measures ANOVA with a Greenhouse-Geiser correction revealed a main effect for stimuli (F=5.14, p= 0.0366) and a main effect for phase (F=4.30, p=0.0464) There was no interaction between phase and stimuli (F=2.65, p=0.0954). Follow up pairwise comparisons showed a significant difference between phase 3 vs phase 2 (t=-3.80, p=0.001) and phase 3 vs phase 1 (t= -4.28, p=0.000).

A follow-up analysis was also conducted with the 12 noticer subjects. A 2 (stimuli: SWS, control) by 3 (phase: 1, 2, 3) repeated measures ANOVA with a Greenhouse-Geiser correction revealed no main effects for stimuli (F=1.08, p= 0.3210) or phase (F=2.94, 0.0846) and no interaction between stimuli and phase (F=0.25, p=0.7377).

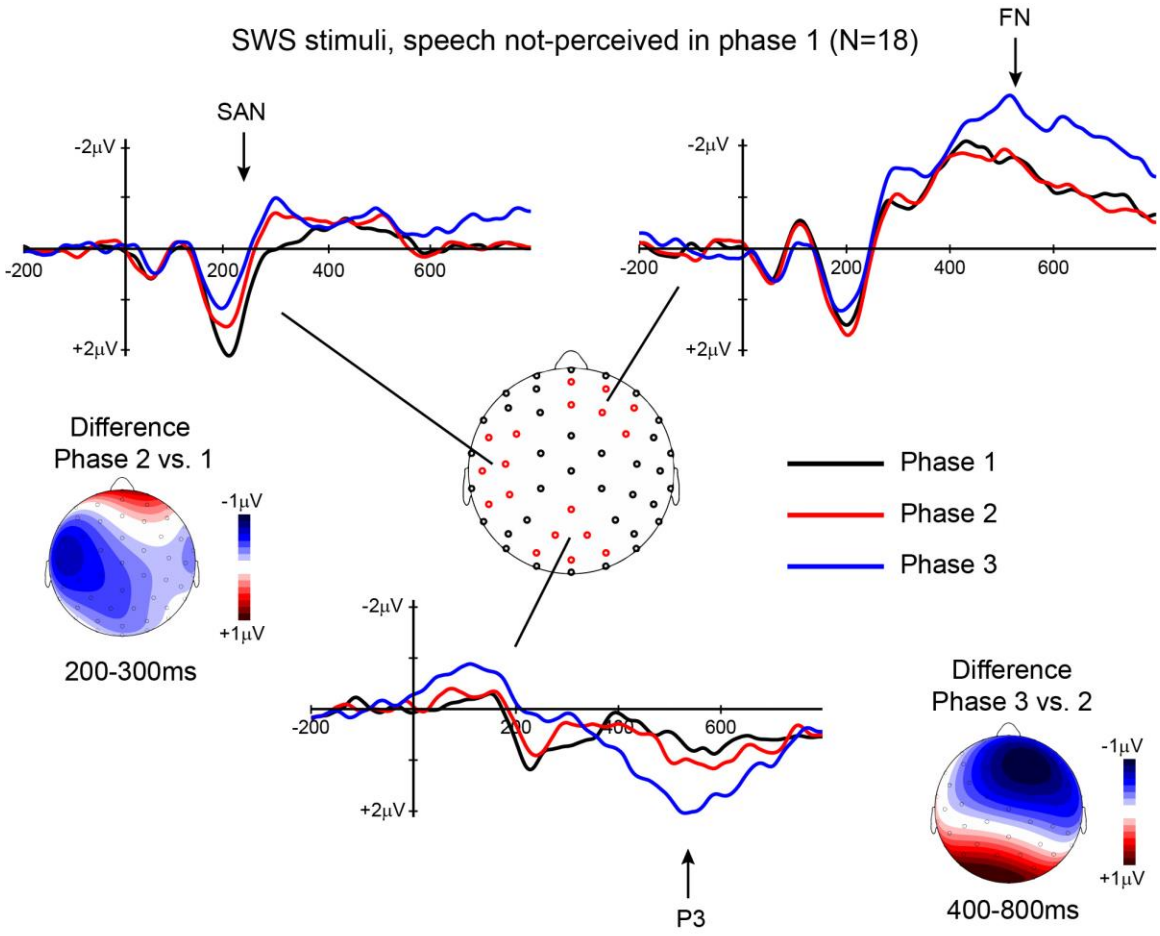


Figure 9: ERPs and Scalp Maps for SWS stimuli:

ERPs of the 18 non-noticer subjects for the three main effects time locked to the beginning of the main speech envelope for SWS stimuli across all three phases. The SAN was measured within a pool of eight left fronto-central electrodes. The SAN refers to the difference that it seen between phases 3 and 3 vs. phase 1 from ~200-300ms (i.e. when speech vs. noise was perceived). A scalp map of the difference between the phase 2 vs. phase 1 at 200-300ms shows the left fronto-central negative distribution of the SAN. The FN was obtained from a pool of eight right frontal electrodes. A scalp map of the difference between phase 3 vs phase 2 from 400-800ms shows the negative right frontal distribution. The P3 was measured in a cluster of eight posterior electrodes and is present from ~400-800ms in phase 3.

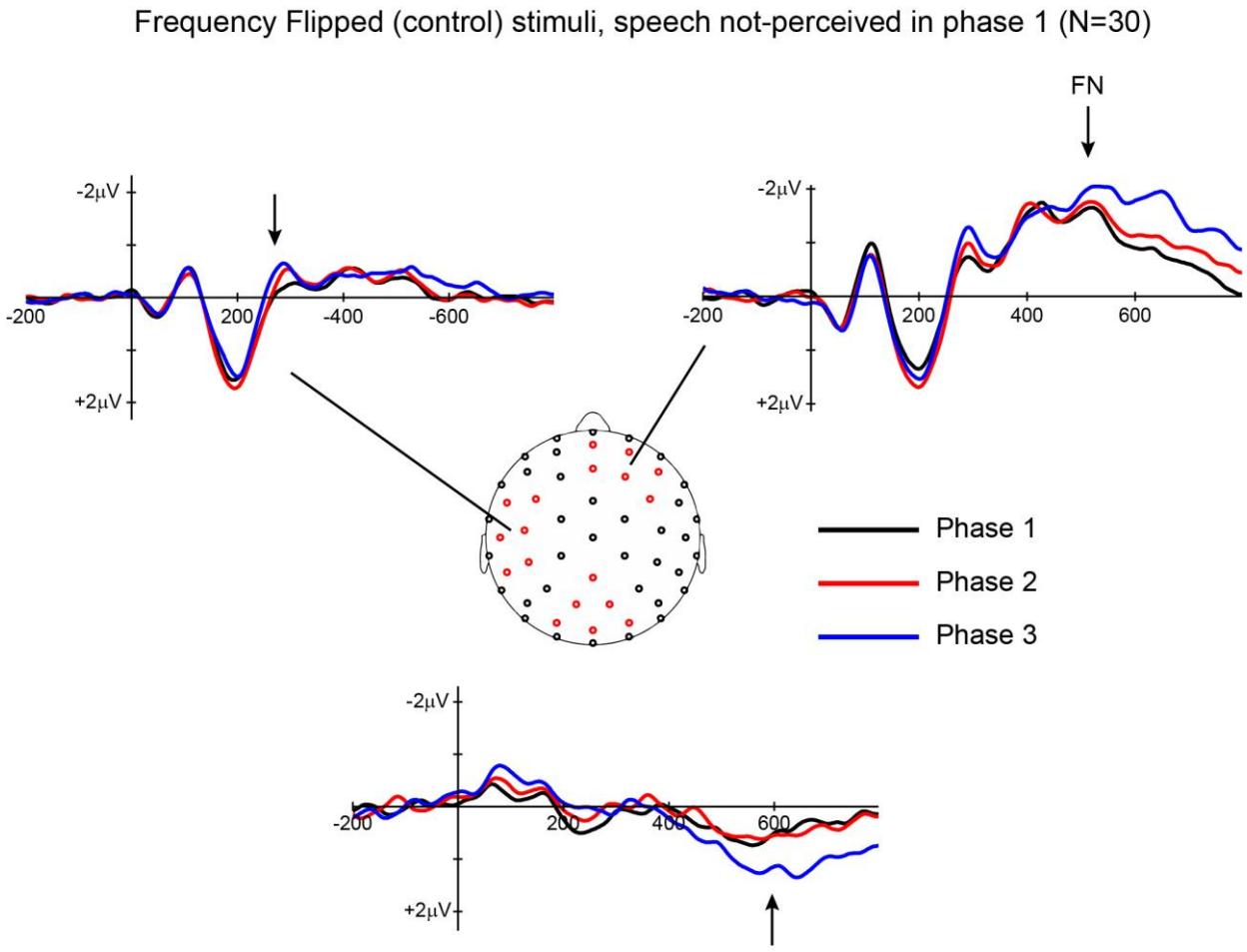


Figure 10: ERPs and for Control stimuli:

ERPs of all subjects (noticers and non-noticers combined) for the three main effects time locked to the beginning of the main speech envelope for control stimuli across all three phases. A pool of eight left fronto-central electrodes shows that there is no SAN present from ~200-300ms unlike with the SWS stimuli. The FN was measured within a pool of eight right frontal electrodes and it still present for these control stimuli at 400-800ms showing a negative right frontal distribution. To see if there is a P3 a pool was made of eight posterior electrodes. A difference is visible in the waveforms at ~400-800ms in phase 3, but was not statistically significant.

## Source Analysis

Low Resolution Electromagnetic Tomography (LORETA) was used to estimate the sources of the ERP effects. The SAN source analysis from 200-300ms showed activation in the superior temporal gyrus, middle frontal gyrus and cuneus (See figure 15). Source analysis of the task effects (P3 and FN) at 400-800ms showed widespread activation across the brain (See Figure 16).

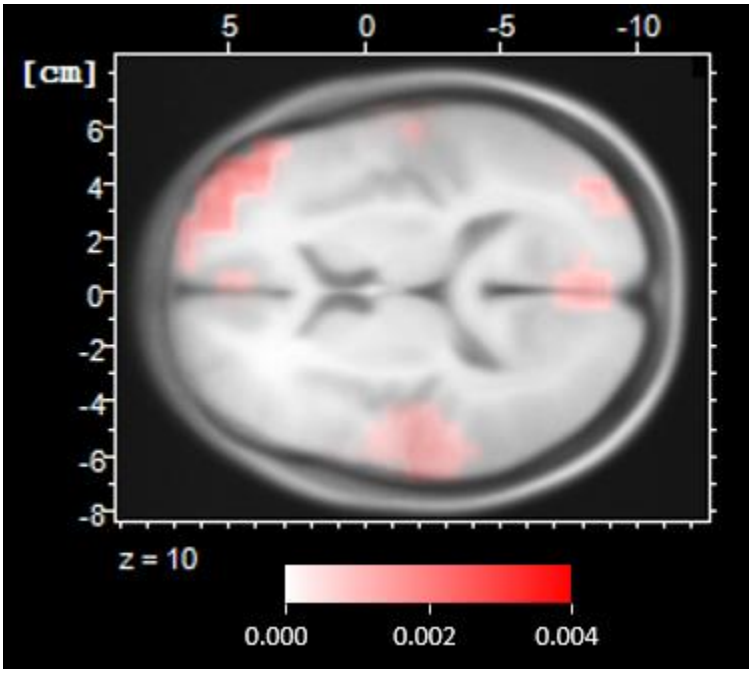


Figure 15: SAN Source Analysis
Low Resolution Electromagnetic Tomography (LORETA) was used to estimate the sources of the SAN ERP effect 200-300ms. It revealed activation in the superior temporal gyrus, middle frontal gyrus and cuneus.

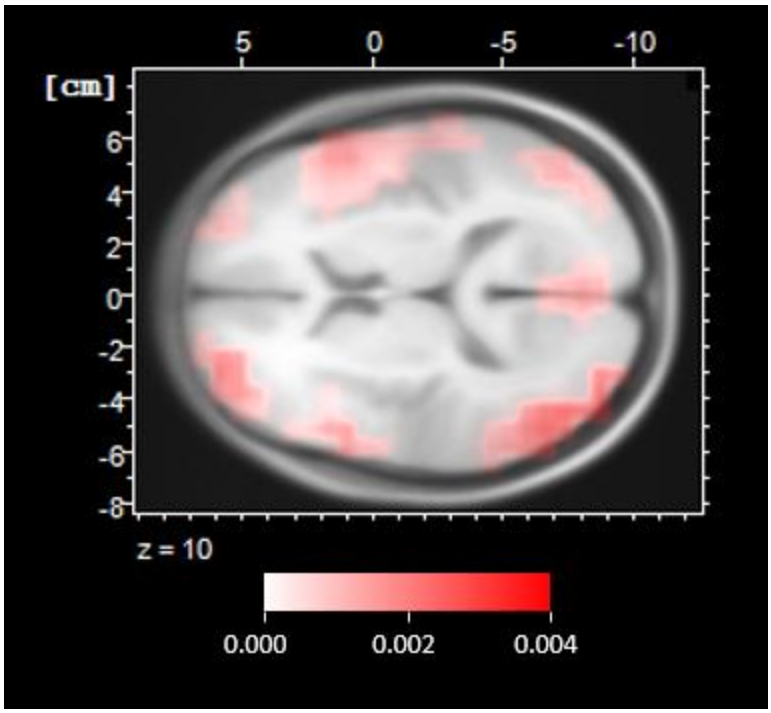


Figure 16: FN and P3 Source Analysis

Low Resolution Electromagnetic Tomography (LORETA) was used to estimate the sources of the P3 and FN ERP effects from 400-800ms that showed widespread activation.

## Time-Frequency Analysis

Six electrodes from the left frontal scalp (7, 19, 20, 35, 36, 53,) were pooled. Wavelets centered at 12.86Hz (range: 7.97-18.50Hz) were extracted and mean amplitudes from 100-200ms. A within subjects t-test was run between SWS stimuli in Phase 2 and SWS stimuli in Phase 1 ($t=-2.83$, $p=0.008$). A within subject t-test was also run for control stimuli in phase 2 vs control stimuli phase 1 ($t= -0.79$, $p=0.432$). This indicates a low beta\high alpha suppression from phase 2 to phase1 in the SWS condition (See figure 11). This is not present in the control condition (See figure 12).

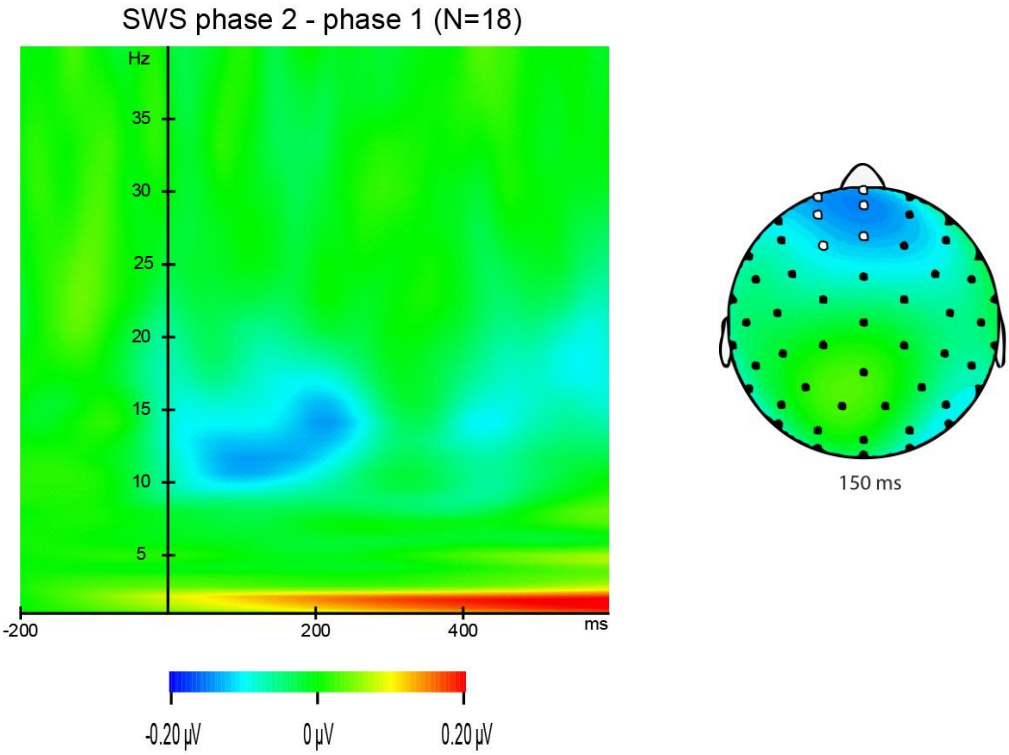


Figure 11: Time frequency SWS phase 2- phase 1

Time-Frequency plot of the difference between SWS stimuli in phase 2 – phase 1 of the eighteen non-noticer subjects. Six frontal electrodes were pooled and are indicated in white on the scalp map. Wavelets centered at 12.86Hz (range: 7.97-18.50Hz) were extracted and mean amplitudes from 100-200ms. There was a negative distribution in the high alpha low beta range. The scalp map shows the distribution of the effect is on the frontal midline at 150ms.

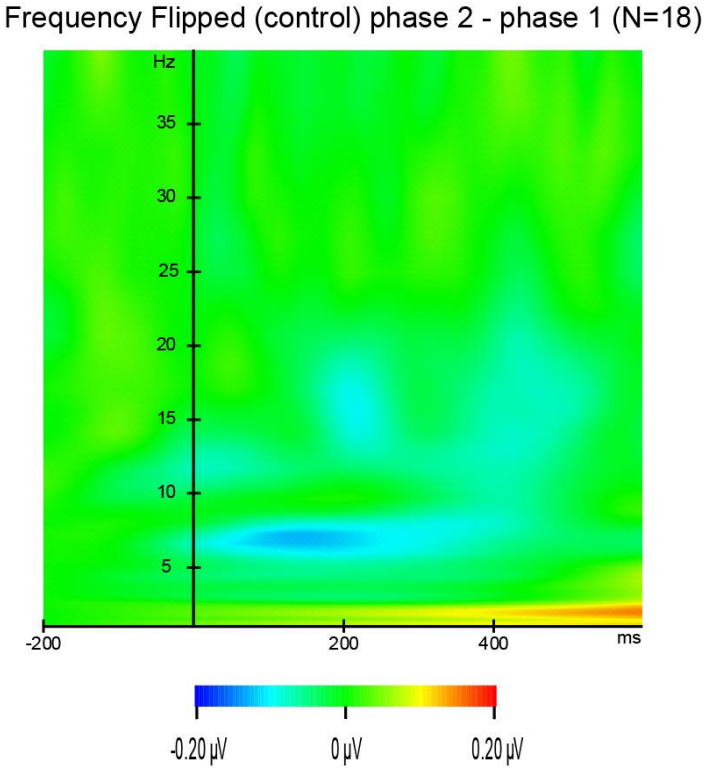


Figure 12: Time frequency SWS phase 2- phase 1

Time-Frequency plot of the difference between control stimuli in phase 2 – phase 1 of the eighteen non-noticer subjects. The same six central frontal electrodes were pooled as indicated in white on the scalp map in Figure 11. Wavelets centered at 12.86Hz (range: 7.97-18.50Hz) were extracted and mean amplitudes from 100-200ms. There was no effect present at this frequency range or time for the control stimuli.

Six electrodes (7, 19, 20, 36, 53, 71) were pooled. Wavelets were extracted at 10.64Hz (range: 6.60-15.31Hz) and the mean amplitude from 150-250ms from a SWS Phase 3- phase 2 difference wave and a Control Phase 3- Phase 2 (See Figure). A within subjects t-test was run between SWS stimuli in Phase 3 and SWS stimuli in Phase 2 (t=-2.09, p=0.044). A within subject t-test was also run for control stimuli in phase 2 vs control stimuli phase 1 (t= -1.43, p=0.163). This indicates an alpha enhancement in between phase 3and 2 in the SWS condition (See Figure 13) which was not present in the control stimuli (See figure 14).

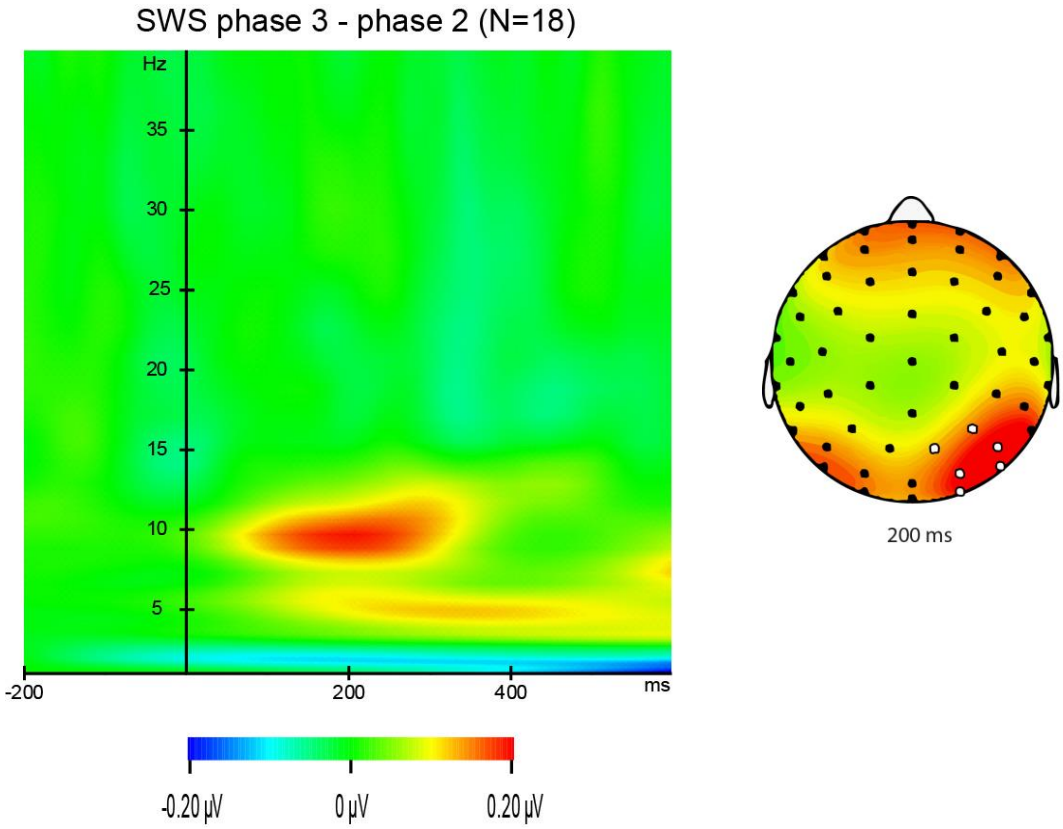


Figure 13: Time frequency SWS phase 3- phase 2

Time-Frequency plot of the difference between SWS stimuli in phase 3 – phase 2 of the eighteen non-noticer subjects. Six right posterior electrodes were pooled and are indicated in white on the scalp map. 10.64Hz (range: 6.60-15.31Hz) and the mean amplitude from 150-250ms. There was a positive distribution in the alpha range. The scalp map shows the distribution a positive effect is right posterior at 150ms.

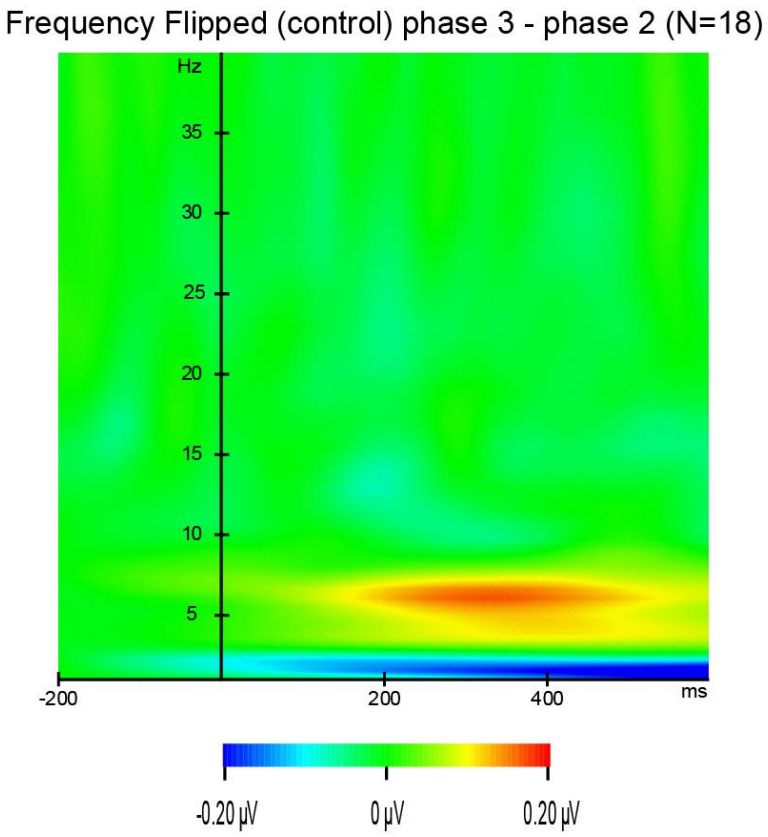


Figure 14: Time frequency control phase 3- phase 2

Time-Frequency plot of the difference between control stimuli in phase 3 – phase 2 of the eighteen non-noticer subjects. Six right posterior electrodes were pooled and are indicated in white on the scalp map. 10.64Hz (range: 6.60-15.31Hz) and the mean amplitude from 150-250ms. There no effect in this frequency or in this time range for the control stimuli.

# Discussion

## Summary of Results

The main goal of this study was to isolate the neural difference associated with perceiving speech versus perceiving noise, while keeping the sensory input and the task constant. We were able to identify a "speech awareness negativity" (SAN) that had a left frontocentral distribution and a 200-300ms latency. The SAN was present in the difference between phase 2 and phase 1 for the SWS stimuli, in the group of subjects who experienced the perceptual shift across these phases. There was no SAN for the control stimuli that were always perceived as noise. There was also no SAN in the group of participants who were aware of the speech content of the SWS in both phase 1 and 2. Time-frequency analysis suggested an additional neural difference between hearing speech vs. noise in SWS: left frontocentral alpha inhibition at ~12 Hz was observed for SWS in phase 2 compared to phase 1. This spectral difference was not present for the contrast of phase 2 to phase 1 in the control stimuli.

Finally, this study allowed for a better isolation of task effects from the perceptual change effects. All previous studies had the SWS speech stimuli as task relevant in the "speech perceived" condition. In the current experiment the SWS stimuli were not task relevant until phase 3. When comparing task relevant vs. task irrelevant SWS, two long latency and widely distributed ERP effects were detected: a P3-like component over the parietal scalp from ~400-800ms, and a sustained frontal negativity from ~400-800ms. Long latency components such as these have previously been suggested to be linked with perceptual awareness, but the current study suggests that they are most likely associated with post-perceptual task-related processing instead.

# Theoretical Implications

These results can be used to further support the idea of the predicting coding hypothesis and may offer some initial insights into the potential neural mechanisms underlying predictive coding in auditory speech perception. Since the sine-wave speech remained the same physically throughout the experiment, and because the stimuli were equally task irrelevant in phase 1 and 2, the neural differences between speech and noise most likely reflect differences in perceptual processing. What could be happening when we first hear the SWS is that we make an incorrect hypothesis about what we are hearing. Once we are given more information (such as hearing the original speech versions of the stimuli during the intervening training), we are then able to update that hypothesis by connecting it to this new knowledge we have on the speech content.

For most of the participants who became aware of the sine-wave speech during phase 1, they did not immediately hear the speech content within the stimuli. Instead their perception changed somewhere during the middle or end of the first phase. These participants were able to update their hypothesis based on fragmentary/ambiguous/noisy sensory input combined with prior general knowledge about speech, without needing further information on the particular speech stimulus (as provided by the intervening training).

The speech awareness negativity (SAN) that was found in this study may index one of the key mechanisms supporting predictive coding in the brain, although it remains unclear whether the SAN is associated with the updated prior, the error minimization process that results from the comparison of the prior with the incoming sensory information, or the resulting perception itself. Future studies may be able to build upon the current findings, by manipulating priors and sensory input separately, matching or mismatching them to various degrees, while measuring the SAN.

Other implications of this study are based on the results of the P3. There has been much debate in the consciousness literature of whether the P3 is a neural correlate of conscious perception (Rutiku, Martin, Bachmann, & Aru, 2015; Salti, Bar-Haim, &

Lamy, 2012). Because the P3 was only found in the current study when the stimuli were task relevant, this suggests that the P3 is not a neural correlate of conscious perception (in this case speech perception) but is instead tied to the task relevancy of the stimuli.

The time-frequency data showed a difference in alpha oscillations across phases. Alpha oscillations have previously been shown to be linked to inhibition as well as to the top- down processes of attention (suppression and selection) (Roach & Mathalon, 2008). There is a negative correlation between alpha power and attention. When there is less alpha power attention is increased. This could indicate that the suppression of alpha in phase 2 vs phase 1 means that more attention is being focused on the SWS speech stimuli when they are perceived as speech than noise. Now the stimuli are not just categories as being noise, but it is categorized as speech and subsequently a specific word. These added steps may be what's causing an increase in attention to the SWS stimuli in phase 2. The difference between phase 3 and phase 2 shows the opposite finding, an increase of alpha and therefore a possible decrease in attention. However, the alpha burst has a different scalp location and is more likely due to a task effect.

# Previous SWS Papers

One of the main goals of experiment was to address some of the limitations of previous sine-wave speech papers. We were able to isolate some of the task effects by showing that the P3 and the FN only showed up in the phase three condition. The component we did isolate to the perceptual switch from speech to noise, the SAN, was then source analyzed. The finding that one of the likely sources of the SAN is the superior temporal gyrus is consistent with the previous sine-wave speech papers; almost all previous studies reported activation in this area. Another possible source was the middle frontal gyrus which is consistent with the Khoshkhoo et al. (2018) study which found some frontal effects.

# Limitations

One of the main limitations of the current study is that the SWS stimuli may not have been truly task irrelevant in phase 2. Since there was an extended training and a recognition task on the SWS immediately before phase 2 participants may have treated these stimuli as relevant in phase 2 (i.e. due to carry-over from the between-phase training and testing session) even when performing the one-back task on the pure tone stimuli. Although in phase 2 we instruct the participants to only complete the one-back task on the tones, participants may have been expecting the SWS stimuli to have some importance, since we had previously drawn a considerable amount of attention to them with the questionnaire, the passive training, and the active recognition test. This could explain the frontal activity we are seeing when doing the source analysis. Another, limitation is there was some effect of phase that was occurring along with the SAN. We can see this by the fact that there was a significant difference between phase 3 and phase 2 for the sine-wave speech stimuli even though perception didn't change between these two phases.

# Future Directions

In order to combat some of these problems, a future study is currently being designed in the SCALP Lab. This study will limit the training session between phase 1 and 2 and eliminate the recognition test. This will draw less attention to and place less emphasis on the SWS stimuli. This study will also counterbalance the order of phase 2 and 3 across participants. This way half the participants will perform a task on the SWS immediately after they learn to perceive the stimuli, while going back to the task on the pure tones in the third phase (possibly helping to reduce the task relevance of these stimuli). This counterbalancing can also help to reduce any order effects that are caused by always running the three phases in the same order.

Another key change that could be made for a future experiment is that in addition to SWS words and control stimuli, SWS pseudowords could also be included (similar to

some of the previous SWS studies summarized above, which used pseudowords or syllables instead of words). The SWS pseudoword stimuli provide an element of speech content without any attached meaning. This will allow us to isolate the difference between perceiving noise, perceiving speech without semantic content, and perceiving speech with semantic content. If there are difference in neural activity between all three of these then it could provide insights relevant for the predictive coding account of speech perception.

While in this lab we are limited by only having access to EEG recording methods, it would be ideal if in the future a SWS studying using the three phase paradigm was run using MEG or fMRI. Although we did source estimation the spatial resolution of EEG is poor. Some of the brain areas estimated to play a role in the current study may change if we used a brain recording method with high spatial resolution.

# Appendix A: Speech Awareness Questionnaire

## Questionnaire

*In the phase of the experiment you just completed, you were asked to focus on the low, middle, and high pitched tones and ignore the computer-generated noises. However, we are also collecting information on how you perceived the computer-generated noises.*

1. In your own words, describe what the computer-generated noises sounded like.

2. Did you hear any of the following in the computer-generated sounds? For each of the categories mentioned in the table below, *please circle only one number representing your experience.*

1=very confident I did not hear it
2= confident I did not hear it
3=uncertain
4=confident I did hear it
5=very confident I did hear it

| | | | | | |
|---|---|---|---|---|---|
| **Distorted music** | 1 | 2 | 3 | 4 | 5 |
| **Distorted words** | 1 | 2 | 3 | 4 | 5 |
| **Distorted environmental sounds** | 1 | 2 | 3 | 4 | 5 |
| **Distorted animal sounds** | 1 | 2 | 3 | 4 | 5 |

3. If you answered “4” or “5” to any of the categories in Question 2, did your perception of the computer-generated sounds change at any point over the course of the task (skip if you circled “3”, “2”, or “1” for all categories)?

YES / NO

If YES, please circle *when* your perception of the sounds changed.

| | | | |
|---|---|---|---|
| During the first quarter | During the second quarter | During the third quarter | During the last quarter |

4. If in Question 2 you responded “4” or “5” to any of the sounds listed, please indicate how often you heard that sound by circling one of the options below (skip any sounds in which you responded 1-3 on the last question).

Distorted music:

| | | | | |
|---|---|---|---|---|
| Less than 10 times | 10-25 times | 25-50 times | 50-100 times | more than 100 times |

Distorted words:

| | | | | |
|---|---|---|---|---|
| Less than 10 times | 10-25 times | 25-50 times | 50-100 times | more than 100 times |

Distorted environmental sounds:

| | | | | |
|---|---|---|---|---|
| Less than 10 times | 10-25 times | 25-50 times | 50-100 times | more than 100 times |

Distorted animal sounds:

| Less than 10 times | 10-25 times | 25-50 times | 50-100 times | more than 100 times |
|---|---|---|---|---|

5. If you marked a “4” or “5” for hearing the computer-generated noises as “distorted words” during the last phase, please write down what words you heard in the computer-generated noises.

# Appendix B: Speech Recall and Recognition Questionnaire

1.During that last phase in which you were responding to repeated tones, you were also presented with some of the sine-wave speech words from your training blocks. Please list all the words you heard during the phase you just complete

2. From this list of words, please select all of the words that you heard during the last experimental block when you were responding to the tones (please do not select words that you only heard in training).

Chair
Yard
Chill
Wave
Bird
World
Brain
Tree
Speech

# Appendix C: Electrode Coordinates Map

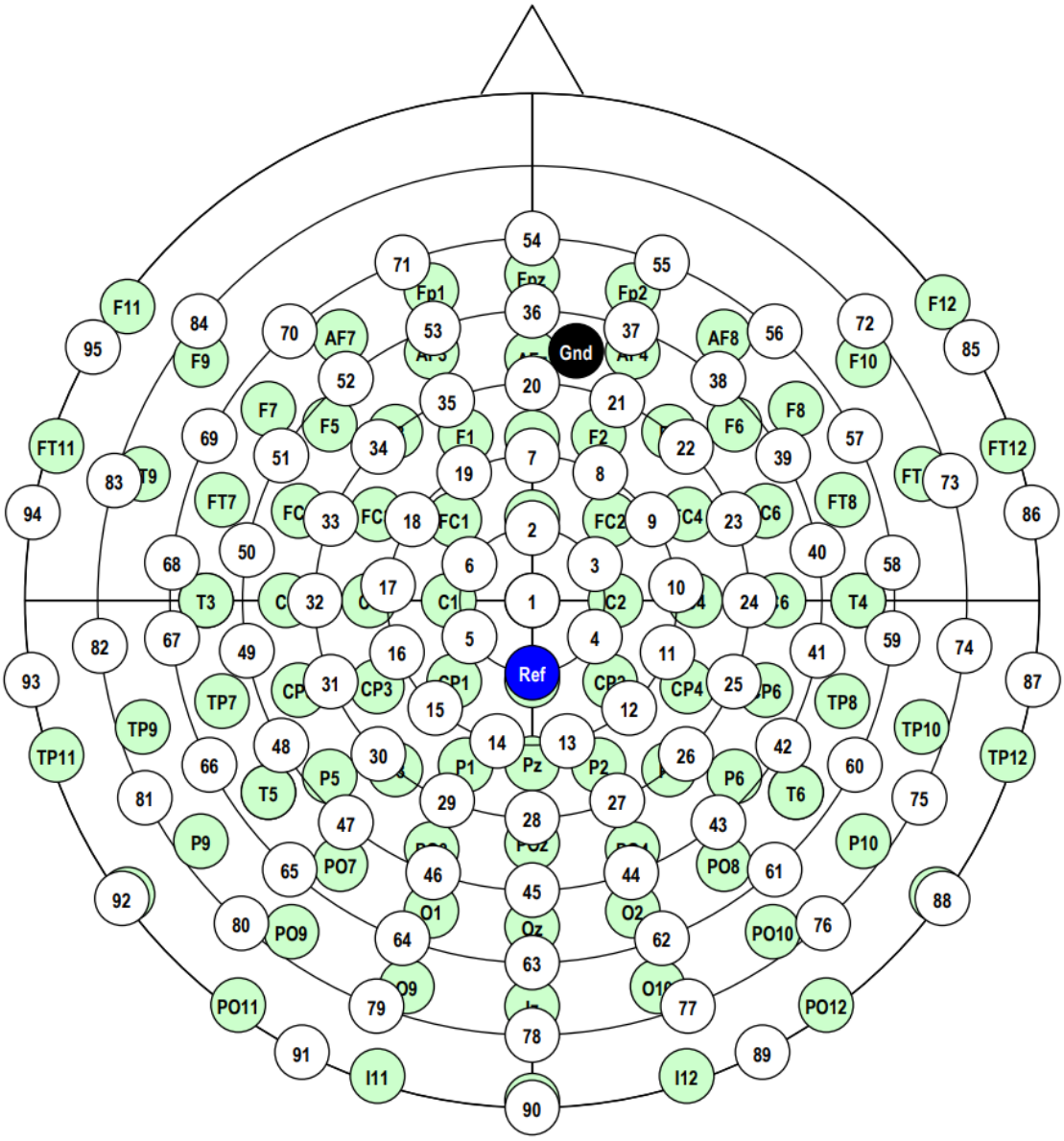


Figure 17: Electrode Coordinate Map

This graph shows equidistant 96Channel Montage No. 60 and underneath the positions of Int. 10%-System. Both are centered around Cz.

# Appendix D: SWS Waveforms and Spectrograms

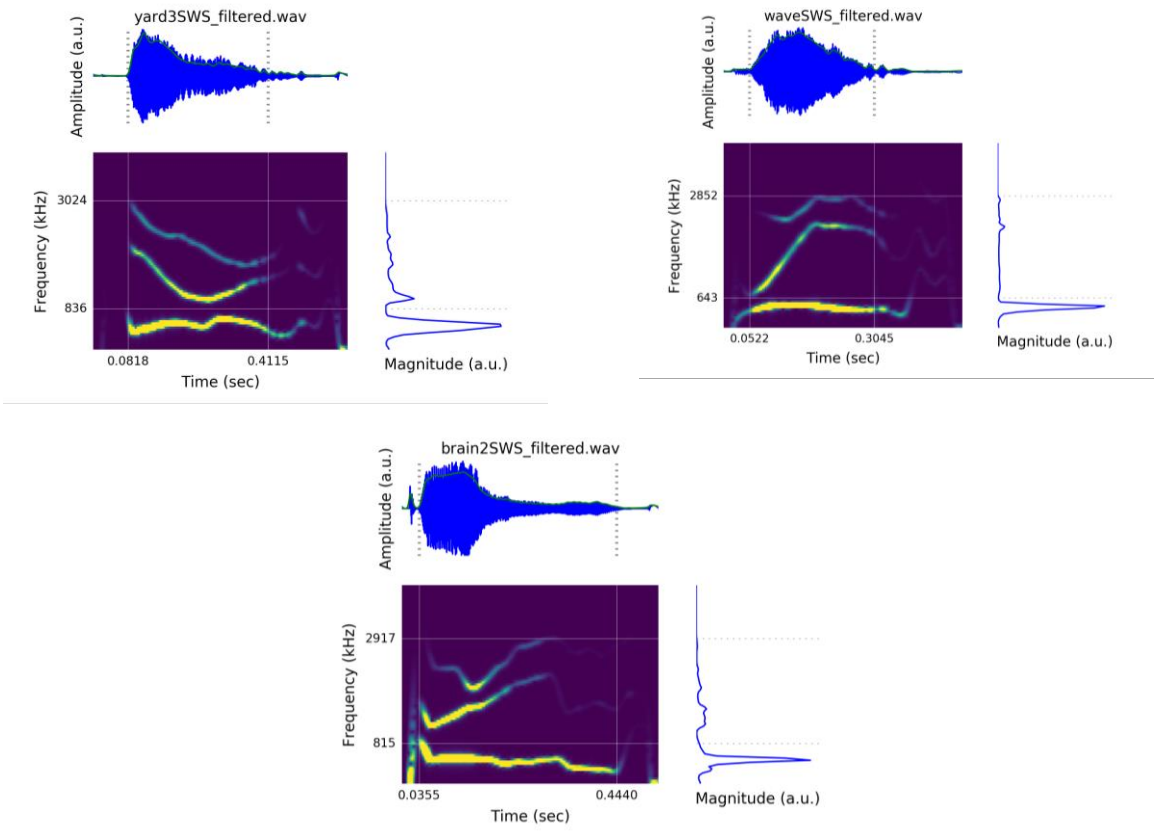


Figure 18: SWS stimuli Spectrograms

The spectrograms and wave forms of the three SWS stimuli. The dotted line on the wave form represent the main speech envelope.

# Bibliography


Aitchison, L., & Lengyel, M. (2017). With or without you: predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, *46*, 219–227. https://doi.org/10.1016/j.conb.2017.08.010

Aru, J., Bachmann, T., Singer, W., & Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neuroscience & Biobehavioral Reviews*, *36*(2), 737–746. https://doi.org/10.1016/j.neubiorev.2011.12.003

Banai, K., & Amitay, S. (2012). Stimulus uncertainty in auditory perceptual learning. *Vision Research*, *61*, 83–88. https://doi.org/10.1016/j.visres.2012.01.009

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International*, *5*.

Brock, J., Brown, C., Boucher, J., & Rippon, G. (2002). *The temporal binding deficit hypothesis of autism* (Vol. 14). https://doi.org/10.1017/S0954579402002018

Cassidy, S. (2002). Speech recognition. *Sydney Australia*, 10–35.

Davis, M. H. (n.d.). An Introduction to Sine-Wave Speech. Retrieved January 31, 2019, from https://www.mrc-cbu.cam.ac.uk/people/matt.davis/sine-wave-speech/

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage*, *24*(1), 21–33. https://doi.org/10.1016/j.neuroimage.2004.09.039

Fridriksson, J., Fillmore, P., Guo, D., & Rorden, C. (2015). Chronic Broca's Aphasia Is Caused by Damage to Broca's and Wernicke's Areas. *Cerebral Cortex*, *25*(12), 4689–4696. https://doi.org/10.1093/cercor/bhu152

Friston, K. (2012). The history of the future of the Bayesian brain. *Neuroimage*, *62–248*(2), 1230–1233. https://doi.org/10.1016/j.neuroimage.2011.10.004

Glover, G. H. (2011). Overview of Functional Magnetic Resonance Imaging. *Neurosurgery Clinics of North America*, *22*(2), 133–139. https://doi.org/10.1016/j.nec.2010.11.001

Goldstein, E. B. (2009). *Sensation and Perception*. Cengage Learning.

Gonzalez, R. (n.d.). A New Optical Illusion Demonstrates How Gullible Our Brains Really Are. Retrieved May 2, 2019, from io9 website: https://io9.gizmodo.com/a-new-optical-illusion-demonstrates-how-gullible-our-br-1579563464

Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *352*(1358), 1121–1127.

Harkness, D. L., & Keshava, A. (2017). *Moving from the What to the How and Where – Bayesian Models and Predictive Processing*. https://doi.org/10.15502/9783958573178

Hartman, K., Peluzzo, A., Shadani, S., Chellquist, I., Weprin, S., Hunt, H., … Altschuler, E. L. (2017). Devising a Method to Study if Wernicke's Aphasia Patients are Aware That They Do Not Comprehend Language or Speak It Understandably. *Journal of Undergraduate Neuroscience Education*, *16*(1), E5–E12.

Hill, N. J., Gupta, D., Brunner, P., Gunduz, A., Adamo, M. A., Ritaccio, A., & Schalk, G. (2012). Recording Human Electrocorticographic (ECoG) Signals for Neuroscientific Research and Real-time Functional Cortical Mapping. *Journal of Visualized Experiments : JoVE*, (64). https://doi.org/10.3791/3993

Jones, M., & Love, B. C. (2011). Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(4), 169–188. https://doi.org/10.1017/S0140525X10003134

Khoshkhoo, S., Leonard, M. K., Mesgarani, N., & Chang, E. F. (2018). Neural correlates of sine-wave speech intelligibility in human frontal and temporal cortex. *Brain and Language*, *187*, 83–91. https://doi.org/10.1016/j.bandl.2018.01.007

Liebenthal, E., Binder, J. R., Piorkowski, R. L., & Remez, R. E. (2003). Short-Term Reorganization of Auditory Analysis Induced by Phonetic Experience. *Journal of*

*Cognitive Neuroscience*, *15*(4), 549–558. https://doi.org/10.1162/089892903321662930

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural Substrates of Phonemic Perception. *Cerebral Cortex*, *15*(10), 1621–1631. https://doi.org/10.1093/cercor/bhi040

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. MIT Press.

Luck, S. J., & Kappenman, E. S. (2011). *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press.

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., & Sams, M. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *NeuroImage*, *30*(2), 563–569. https://doi.org/10.1016/j.neuroimage.2005.10.002

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science (New York, N.Y.)*, *212*(4497), 947–949.

Roach, B. J., & Mathalon, D. H. (2008). Event-Related EEG Time-Frequency Analysis: An Overview of Measures and An Analysis of Early Gamma Band Phase Locking in Schizophrenia. *Schizophrenia Bulletin*, *34*(5), 907–926. https://doi.org/10.1093/schbul/sbn093

Rutiku, R., Martin, M., Bachmann, T., & Aru, J. (2015). Does the P300 reflect conscious perception or its consequences? *Neuroscience*, *298*, 180–189. https://doi.org/10.1016/j.neuroscience.2015.04.029

Salti, M., Bar-Haim, Y., & Lamy, D. (2012). The P3 component of the ERP reflects conscious perception, not confidence. *Consciousness and Cognition*, *21*, 961–968. https://doi.org/10.1016/j.concog.2012.01.012

Samuel, A. G. (2010). Speech Perception. *Annual Review of Psychology*, *62*(1), 49–72. https://doi.org/10.1146/annurev.psych.121208.131643

Singh, S. P. (2014). Magnetoencephalography: Basic principles. *Annals of Indian Academy of Neurology*, *17*(Suppl 1), S107–S112. https://doi.org/10.4103/0972-2327.128676

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

The Bayesian Brain: An Introduction to Predictive Processing. (2018, July 28). Retrieved April 30, 2019, from Mindcoolness website: https://www.mindcoolness.com/blog/bayesian-brain-predictive-processing/

Time frequency tutorial - SCCN. (n.d.). Retrieved April 19, 2019, from https://sccn.ucsd.edu/wiki/Time_frequency_tutorial

Tivadar, R. I., & Murray, M. M. (2019). A Primer on Electroencephalography and Event-Related Potentials for Organizational Neuroscience. *Organizational Research Methods*, *22*(1), 69–94. https://doi.org/10.1177/1094428118804657