University of North Carolina Asheville
Journal of Undergraduate Research
Asheville, North Carolina
Fall 2023
David Bortolotto

The Neural Correlates of Perceptual Learning of Sine-Wave Speech

Psychology
The University of North Carolina Asheville
One University Heights
Asheville, North Carolina 28804 USA
David Bortolotto
Faculty Mentor(s): Dr. Michael Neelon

Abstract

The present study investigates the changes in electroencephalogram (EEG) responses that occur when listeners learn to perceive sine-wave speech (SWS) as clearly intelligible speech. SWS is an artificial acoustic signal made by reducing the complex amplitude and frequency changes of natural speech to several time-varying sinusoids (Remez, 2008). Typically, listeners begin to hear speech in SWS after exposure to the original audio recording on which it is based. Our study attempts to extend the reported finding of an event-related potential (ERP) called the “Perceptual Awareness Negativity” (PAN) that may occur when listeners start hearing SWS as speech rather than as unidentifiable electronic noise (Zhu et al, *under review*). At the start of each session, participants listened to SWS, pure-tones, and spectrally rotated control words and responded when they heard any sounds repeat. Following this, participants completed a speech training phase in which SWS and natural speech tokens were paired together. After speech training participants repeated the initial phase yet this time with updated speech awareness. Behavioral results show that 80% (25/31) of participants did not hear speech content in the first phase of exposure to SWS, but reported hearing SWS as intelligible words after training. These results confirmed the perceptual experience of learning to hear SWS as speech after proper exposure. ERPs were analyzed for the presence of a left-lateralized fronto-central negativity in the 200-300 ms post-stimulus time range similar to that of the PAN. A negative shift in the grand-averaged ERP was observed in participants hearing SWS as speech after speech training between blocks 1 and 2 (i.e., non-noticers). At the same time, no similar negative shift was observed across blocks for these same listeners in the ERPs to spectrally-rotated controls, indicating the absence of a PAN for stimuli not perceived as speech.

Keywords: speech perception, perceptual learning, perceptual insight, sine-wave speech, event-related potential (ERP)

Background

Humans have a remarkable ability to make sense of ambiguous incoming sensory information, regardless of stimulus modality (Jean-Luc Schwartz, et al., 2012). Examples of the perceptual restoration of degraded sense data across multiple sensory systems suggest that humans have gained adaptive means for discerning sensory signals amidst noise in their environment. In the realm of visual perception, humans often recognize complex patterns in visually ambiguous, disorganized, or physically incomplete stimuli, such as blurry faces (Sinha, 2002), camouflaged animals, or when seeing spatial objects from a novel angle (Cohen, 2015). This ability has also been found in other senses, such as olfactory (Millar, 2017), gustatory (Sanchez, Dwyer, Honey, et al., 2022), tactile (Rodríguez & Angulo, 2014), and multi-sensory (Beer, Batson & Watanabe, 2011) modalities.

The capacity to identify meaning in ambiguous stimuli extends to audition, such as when listeners perceptually restore phonemic and semantic content in physically impoverished sounds (Sohoglu & Davis, 2020; Warren, 1971). Speech perception in particular illustrates well how listeners can successfully identify target sounds despite impoverished input, whether due to competing environmental sounds, individual variations across speakers, or against a listener's own contradicting top-down influences (Leibold, 2017; Wang & Xu, 2021; Wong, Uppunda, Parrish, et al., 2008). A particularly powerful example of listeners' abilities to successfully perceive degraded speech is "sine-wave speech" (SWS). Sine-wave speech is an artificial acoustical signal constructed by representing the complex amplitude and frequency of speech formants using 3-4 time-varying sinusoidal pure tones (Remez, 2008). What makes sine-wave speech an effective tool for exploring the auditory enhancement of degraded sounds is the tendency to initially perceive a SWS replica of a natural word as an unrecognizable, artificial sound, then hearing it as intelligible human speech after a short period of exposure to the original natural speech sample (Mottonen, 2006; Sheffert, Pisoni, Fellows & Remez, 2002). Sine-wave speech thus allows researchers to investigate the internal subjective experience of sensory input (i.e., perception) without altering the underlying physical characteristics of the stimulus itself.

Previous Research

Perceptual Learning

The experience-dependent changes which account for comprehension of clear speech in SWS stimuli is a form of auditory perceptual learning (Gold & Watanabe, 2010). Perceptual learning relies on prior knowledge to discriminate between sensory data, with effects that are generally considered short-term or long-lasting (Watanabe & Sasaki, 2015*)*, and is strongly associated with increased sensitivity to originally diminished or degraded stimuli. Further, perceptual learning is distinguished from other

forms of learning in that training-induced exposure to a target stimulus alone can be enough to create a lasting change in perception (Seitz, 2017).

Perceptual Insight

Auditory perceptual learning of SWS after training points to a sudden enhancement of speech comprehension (Mottonen, 2006; Liebenthal, Ellingson, Spanaki, et al., 2003) indicating a co-occurring process associated with perceptual learning, often called "perceptual insight" (Ruben, Nakayama, & Shapley, 2002). Perceptual insight is marked by sudden changes in performance while evaluating, discerning, or discriminating between stimuli, which are sometimes called "all-or-none" events (Sekar, Findley, Llinás, 2012). Ruben, Nakayama, and Shapley (2002) used a visual analogy to capture perceptual insight by altering a plain image of a tree frog through manual two-toning and application of a Gaussian filter. Subjects were shown the artificially degraded image and were initially unable to recognize the tree frog. However, after receiving a clue from researchers about the image's identity or by being directly exposed to the original picture, subjects experienced rapid shifts in perception, at once seeing a tree frog in the degraded image.

Previous literature on SWS has shown a similar effect, where subjects are initially unable to perceive meaningful speech content, but after brief exposure to the source recording, experience a sudden reversal in perception and hear SWS as intelligible spoken words. This change in perception is unlike that of bistable or multistable stimuli such as the Necker cube or face-vase illusion. In both, subjects can switch between two perceptual experiences, similar to the "Yanny or Laurel" auditory illusion, where people are able to switch attention between higher and lower frequency bands in order to hear different words (Bosker, 2018). Rather, the perceptual learning of sine-wave speech is an example of "one-shot learning" (Lake, Salakhutdinov, Gross, & Tenenbaum, 2011): once subjects are exposed to the natural speech versions of SWS stimuli they are typically unable to "unhear" speech content in the formally ambiguous sounds (Remez, 2008).

Neural correlates of Auditory Perceptual Learning

Research is limited regarding the neural correlates of auditory perceptual learning of artificially degraded speech. Previous research has used the following model: subjects are presented with artificially degraded speech attributes in a training-independent phase followed by a brief speech exposure/training period. Once subjects have been exposed to the undegraded natural speech versions of SWS stimuli, they are given the same task as in the training-independent phase, but this time with updated speech awareness. Researchers have used both EEG and fMRI to record brain changes while subjects undergo the aforementioned procedure.

Libenthal et al. (2003) focused on neural mechanisms for speech perception using both EEG and fMRI. Liebenthal et al. observed listeners as they completed an auditory discrimination task in which participants were trained to perceive a nonspeech two-tone complex word as speech. A voltage distribution across the scalp of subjects of

mismatch negativity ERP's showed a slight negativity among fronto-central electrodes including those associated with the left Hershel gyrus (i.e., primary auditory cortex). fMRI results showed enhanced neural activation in response to speech training in the left Hershel gyrus and left superior temporal gyrus (STG).

In a similar auditory discrimination paradigm, Dehaene-Lambertz et. al. (2005) used simultaneous fMRI and EEG recording in order to monitor the time course of sudden speech perception and to locate definitive neural correlates of speech versus non-speech perception. Mismatch negativity ERPs showed an earlier onset time in response to phonemic vs non-phonemic changes, indicating a coding preference for speech over non-speech sounds. fMRI results indicated greater neural activation in the left superior temporal gyrus and sulcus in conditions in which subjects perceived speech.

Mottonen et al., (2006) conducted an fMRI study using SWS aimed at understanding where in the brain speech versus nonspeech acoustical signals are processed. Mottonen et al scanned listeners as they classified SWS versions of nonsense words in a pre-speech condition, followed by another scan after listeners had been trained to hear the same stimuli as speech. They found that the stimuli produced stronger activation in the left posterior superior temporal sulcus (STS) when perceived as speech.

Hendry (*2019*)*, Zhu et al., (under review)*

Two recent studies have most informed the current research. Both used EEG to investigate the time course and neural correlates of speech versus nonspeech perception using SWS. Hendry (2019) aimed to isolate the neural correlates of speech learning separately from task relevance, which may show attentional-related changes. Using a 3-phase design Hendry had subjects perform a series of one-back tasks on pure-tone stimuli in an initial block, which also included SWS and spectrally rotated control sounds in the stream of ongoing stimuli. This phase was followed by a speech awareness questionnaire, then subjects were exposed to the natural recordings underlying the sine-wave speech stimuli presented earlier, before completing another phase (again responding to pure tone repetitions) of the experiment with updated speech awareness. This was followed by a final phase in which subjects performed one-back tasks on the SWS stimuli instead of the pure-tones. In their analysis, Hendry averaged over an a-priori region of interest of left frontocentral electrodes to generate ERPs, revealing a negative shift in the waveforms between Phase 1 and 2 within a 200-300 ms window after token onset. Hendry called this ERP shift the "Speech Awareness Negativity" (SAN).

Zhu et al (*under review*) replicated Hendry's approach by also accounting for task effects while listening. Their EEG study also presented in an initial block a stream of sine-wave speech words along with modified spectrally-rotated versions and pure tones, with listeners required to respond to any one-back repeats of the tones. This block was followed with a speech training sequence of sine-wave speech followed by their natural versions. After training, Zhu and colleagues presented 2 more blocks of stimuli where listeners responded to pure-tone or SWS repeats, respectively, as in Hendry's original study. ERPs were analyzed using an unsupervised multiple univariate statistical

approach which also revealed a negative shift in the 200-300 ms post-stimulus time window. The authors believed this shift to be specifically related to conscious speech perception, and they labeled it as the “Perceptual Awareness Negativity” (PAN) .

Present Study

The purpose of the present EEG study was to measure the perceptual learning of sine-wave speech, with the goal of locating a specific ERP component for speech awareness. A secondary goal was to understand what specific words and their respective linguistic attributes (e.g phoneme types, syllabic count, etc) make ideal one-shot stimuli when converted to sine-wave speech. Finally, while our study borrowed largely from the work of Zhu et al, we also explored the following simplified design to determine if the PAN may emerge under potentially more practical conditions than that used by Zhu and colleagues. First, we reduced the number of stimuli from 9 (3 exemplars each of SWS, spectrally-rotated controls and pure tone) to 6 (2 exemplars of each type). Second, we reduced the number of blocks from 3 (pure-tone attention pre-speech training; pure-tone attention post-training; SWS attention post-training) to 2 blocks during which listeners attend for any repeated stimuli regardless of type (SWS, SR, or pure-tone) both before and after speech training. The EEG portion of our design was meant to replicate the same ERP component identified as the “Perceptual Awareness Negativity” (PAN).

Method

Pilot Study

A pilot study was first conducted before the main experiment as an attempt to systematically vary source words to see which proved most effective in the perceptual learning of SWS. Previous studies have not provided principled explanations for the words used to generate SWS tokens, with some being monosyllabic and others nonsense syllables (Hendry, 2019; Zhu et al., *under review;* Mottonen et al., 2006). We performed a general language analysis in order to explore speech attributes that might lead to the largest gain in learning as measured by the greatest difference between perceived speech ratings before and after training. We chose words from an analysis of word-frequency and conceptual difficulty performed by Rudell (1993) as rated by 24 adult judges (aged 24-57). The words Rudell (1993) tested were taken from a previous computational analysis of American English providing word-frequency comparisons measured in occurrences per million (Kucera & Francis, 1967). We chose words based on low and moderate levels of rated conceptual difficulty, moderate frequency, and chose primarily monophthongal, as well as some diphthongal words. In addition, we selected stimuli based on the presence and number of unvoiced consonants (which should result in gaps in SWS as there are no formants to replace with sinusoids), and number of syllables.

Participants

Thirty-nine participants were included in the pilot study (aged ~18-45). All participants completed the survey virtually. The survey was created and administered using PsyToolKit (Stoet, 2010). Subjects were recruited anonymously on the internet via social media platforms (e.g., Facebook and Reddit).

Stimuli

Hendry (2019) and Zhu et al. (*under review*) selected the words 'brain', 'wave', and 'yard' for their sine-wave speech tokens. They did not specify their selection process and we know of no study that has systematically explored which words are most effective in perceptual learning of SWS. Our a priori considerations for word selection were that natural tokens should be: ideally monosyllabic, as two-syllable words might 1) increase initially hearing SWS as speech due to prosodic cues, and 2) impact ERP formation due to multiple changes in the auditory envelope; monophthongal; and fairly common in the English lexicon. We did, however, chose to include some diphthongal words as well in order to compare their rating effects for participants and to test the differences in format tracking when converted to SWS. 61 candidate words were selected based on their perceived conceptual difficulty and word-frequency in the English lexicon from an analysis performed by Rudell (1993). The words we chose ranged between high difficulty with medium-low frequency to low difficulty with high-frequency in order to test wide differences between these attributes in our pilot study. The words selected along with mean and standard deviation ratings of the frequency and difficulty are listed in Appendix A.

Sine-wave speech versions of each word sound file were created using a script written by Chris Darwin (2003) for the Praat auditory software environment (Boersma, 2001). Natural speech tokens were created using TextoSpeech.io to produce an unidentifiable female speaker that was consistent in delivery across all speech tokens, then imported into Praat software for SWS construction. A small proportion of the SWS stimuli (~10-20%) were "spectrally rotated" using another Praat script by Darwin (2003), which inverts the sinusoids around a center frequency, creating completely unintelligible stimuli to be used as controls for our study (Mottonen et al., 2006; Bent, Loebach, Phillips, & Pisoni, 2011). These spectrally rotated stimuli were chosen due to the fact they contain the same spectral and acoustic content of their SWS counterparts but are otherwise unintelligible and can be used as "filler" stimuli to reduce spontaneous learning of SWS until after training exposure.

Procedure

Before starting the study, respondents were given a brief description of the procedure and then asked for their informed consent. Listeners were asked to wear headphones for the experiment. Across blocks, participants saw an audio playback button on their screen accompanied by the following statement: "How confident are you that you can identify what the sound is." Participants listened to the sound and then rated their confidence by selecting with their mouse a point on an accompanying 5-point Likert scale: "not confident at all, slightly confident, moderately confident, confident, very confident." After responding, a button below the Likert scale allowed participants to progress to the next trial. After completing 40 trials in block one, participants were informed they would begin block two in which an original speech recording would be

presented on each trial followed by its sine-wave speech replica for an additional 40 trials. Participants were asked again to listen to each audio recording back-to-back and then to rate their confidence hearing speech in SWS words. The average duration for participation was *M* = 10.7 minutes. After the study concluded, participants were debriefed on a separate landing page.

Analysis

Seventy-four percent (37/50) of participants completed all trials of the study. Participants who skipped responses, or did not complete both blocks of the survey, or who claimed to hear one of the spectrally-rotated control words as intelligible speech were rejected from analysis. Average ratings across participants for each stimulus and differences in average ratings across survey blocks (non-speech and speech modes) were calculated in Excel.

Results

Average confidence ratings increased between blocks 1 and 2 (*M* = 1.73, *SD* = 0.32) across all stimuli. For the subsequent EEG study, we selected the following words with the largest mean confidence ratings differences (in parentheses) before and after speech awareness training: event (2.1), being (1.1), hence (1.3), short (2.1), small (2), music (2.2), after (1.8), and basic (2.0). While ‘being’ and ‘hence’ saw relatively small differences in confidence ratings across blocks, they were included as practice stimuli words since the practice block was meant to serve as training for the main experiment and not an actual measure of perceptual learning. The words ‘basic’ and ‘short’ were selected as words to be used in the main blocks of the study. For the speech training blocks, ‘small’, ‘after’, ‘music’, and ‘event’ were added along with ‘basic’ and ‘short’ to provide enough exposure to SWS stimuli to ensure perceptual learning.

EEG Study

The goal of the EEG study was to measure the neural correlates of perceptual learning of sine-wave speech, with the specific goal of replicating the distinctive speech-related ERP the “Perceptual Awareness Negativity” as reported by Hendry (2019) and in Zhu et al., (*under review*).

Participants

31 undergraduate participants (age range 18 - 40; 28 right-handed, 3 left-handed) with reported normal hearing took part in the study. Participants received minor course credits for joining the study. Data collection took place in the Neurolab of the Psychology Department at UNC-Asheville.

Stimuli

10 sound stimuli were created using Praat audio software: 8 sine-wave speech replicas of natural speech (basic, short, small, after, music, event, being, hence) along with spectrally rotated (SR) (Darwin, 2003) versions (basic, short, being, hence), to act as speech controls during EEG recording blocks, and 2 pure-tone sounds of 440 Hz and 1000 Hz. Each sound stimulus was edited to a duration of 600ms. Spectrograms for the

different versions of ‘Basic’ and ‘Short' are seen in Fig. 1 and 2. ‘Basic’ natural token (Fig. 1a), sine-wave speech stimulus (Fig. 1b), and spectrally rotated stimulus (Fig. 1c); ‘Short’ natural token (Fig. 2a), sine-wave speech stimulus (Fig. 2b), and spectrally rotated stimulus (Fig. 1c). Observation of the spectrally rotated version of ‘Basic’ shows incorrect “flipping” of sinusoids around a center frequency due to a technical error in script parameters found after completion of the experiment. Possible implications of this are addressed in the limitations section of the present study.

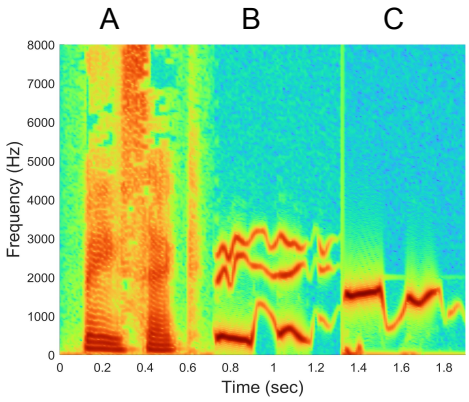

Fig. 1 ‘Basic’ spectrograms

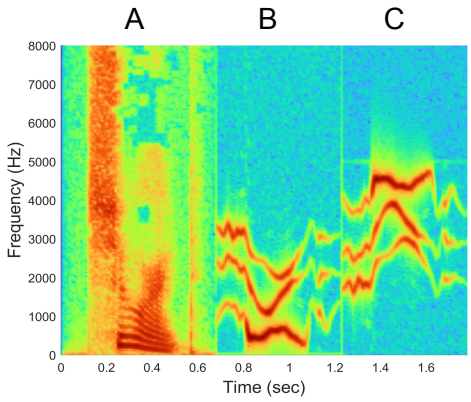

Fig. 2 ‘Short’ spectrograms

Procedure

Before the session began researchers explained that participants were about to take part in an EEG study measuring how humans perceive electronically processed sounds. Participants’ electroencephalograms (EEG) were recorded using a 64-channel BioSemi ActiveTwo acquisition system in conjunction with BioSemi ActiView software (Cortech-Solutions, Wilmington, NC). Participants wore Beyerdynamic DT 990 over-ear headphones and responded to stimuli using a Jelly Comb wireless numeric pad. Participants were asked to sign the informed consent forms for the present protocol, the EEG protocol and its associated HIPAA form.

Participants remained seated at a computer in a sound-attenuated booth (WhisperRoom). In the practice block and Blocks 1 and 2, each trial began with a fixation cross appearing on screen followed by one of the six stimuli (2 SWS exemplars, 2 spectrally rotated controls, 2 pure tones) each presented 50 times in random order with a mean ISI of 800ms (+/- 200ms uniform random jitter) over headphones at a comfortable listening level (~70dB). ~15% of trials in each block contained a repeated stimulus (of any type), determined randomly by software at the start of each block. Listeners were instructed to listen for any repeated stimuli, which they indicated by pressing the “enter” button on the wireless numeric pad.

After completing Block 1, participants filled out a speech awareness questionnaire meant to gauge how confidently they heard speech in any of the sounds. Following Zhu et al we asked subjects to rate their confidence for identifying any of the sounds they heard as distorted words. Ratings ranged according to the following scale:

(1) very confident I did not hear it, (2) confident I did not hear it, (3) uncertain, (4) confident I did hear it, and (5) very confident I did hear it. Participants were then asked to write down any words they heard on a piece of paper (if they heard no words they were prompted to write, “None”). Results from this pre-speech assessment were used to distinguish between subjects who did not hear the sounds initially as speech (henceforth, “non-noticers”) versus those who spontaneously learned to recognize SWS as speech during block 1 (henceforth, “noticers”).

After completing this questionnaire (and regardless of what listeners reported), speech training took place in which participants were told that some of the words they heard were degraded versions of natural speech. Participants were trained on the 2 SWS tokens that were used in Block 1 of the study (‘basic’ and ‘short’), as well as 4 other words (‘music’, ‘event’, ‘exist’ and ‘small’) and their SWS counterparts in order to increase learning exposure. Participants listened to the 6 SWS and original words in the following sequences: SWS →original →SWS, with a 500 ms ISI between each component. Participants listened to 4 repeats of each 3-token sequence for all 6 words. After completion of speech training, participants performed a 6 alternative forced-choice listening-response task on the previous SWS stimuli to confirm they heard the sounds as speech. During the listening-response task, participants listened to one of the 6 SWS tokens used in the training block and had to match the sound to a list of the 6 possible words (the 2 primary SWS words from block 1 along with the 4 additional words).

After SWS training, Block 2 presented the exact same procedure used as in Block 1 except that participants were expected to hear the SWS words as speech sounds. After Block 2 participants were given the same speech awareness questionnaire as done after Block 1. In total, the listening portion of the experiment lasted approximately ~30 minutes, while the entire session for each participant (including EEG prep & clean up) lasted ~1 hour total.

EEG Analysis

Raw recordings were re-referenced to a common average reference, bandpass filtered between 1-40Hz, and ocular artifacts (e.g., eyeblinks) were removed using the EMSE EEG analysis software (Cortech Solutions, Wilmington, NC USA). For all subjects, individual EEG epochs for each stimulus presentation were extracted between -200ms pre- to 800ms post-stimulus and formed into ERPs by averaging all epochs of each stimulus type separately for Blocks 1 and 2. In the case of a small number of participants with extreme noise in any channels, electrode interpolation was performed in order to produce higher-fidelity data for creation of ERPs.

PAN extraction

Following the procedure first reported in Hendry, the PAN was extracted for each subject by further averaging ERP voltage values across the 200-300 ms post-stimulus time window, pooled across left fronto-central electrodes (C1, CP3, C3, FC3, CP5, C5, and FC5 in the Biosemi 64-channel montage). PAN values for each subject were averaged one last time across the 2 exemplars of each stimulus type and submitted to a 2 (Block 1 vs Block 2) x 3 (SWS, spectrally rotated, pure tones) analysis of variance.

Results

Behavioral Results

25 participants (25/31) reported “none” when prompted to write down any words they might have heard in Block 1 prior to speech training. 6 listeners (6/31) reported hearing at least one English word in the same context of the experiment, with 3 of 6 (3/31) recording at least one of the target SWS tokens before receiving speech training. As with Hendry (2019) and Zhu et al., (*under review*), we separated participants into two groups for later analysis: “noticers” (those who reported hearing words in Block 1) and “non-noticers” (those who do not). In addition, one listener was reported as a non-learner and excluded from further analysis. Our criteria for exclusion for this non-learners were the following: no change between confidence ratings in Block 1 and Block 2, lower than 90% on listening-response task (no participants other than the non-learner scored lower than 90% on the listening-response task), and did not report both target SWS words (‘basic’ and ‘short’) in the questionnaire posed at the end of Block 2 (after speech training).

Repeated-measures ANOVA were performed on the confidence ratings across blocks separately for noticer and non-noticer groups. The non-noticer group showed a significant increase in confidence ratings after speech training ($F(1, 22) = 129.8$, $p < .001$, $\eta^2_p = 0.855$), while there was no significant change in ratings for the noticer group ($F(1, 5) = 2.5$, p = n.s.). These results are shown in Figures 3 and 4.

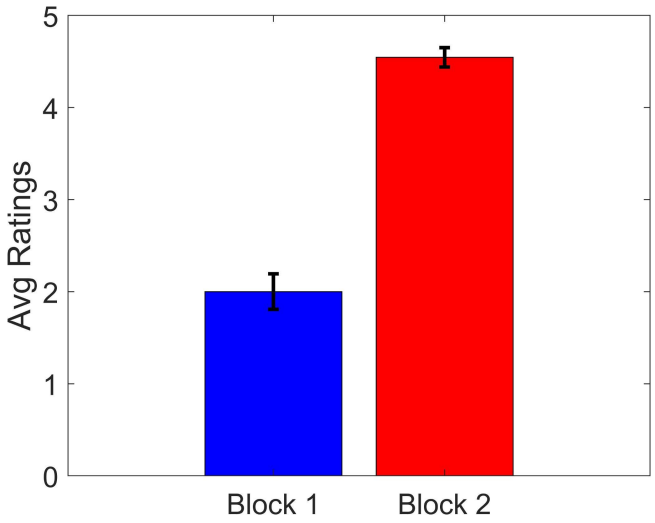

Fig. 3 Non-noticer group participants showed a 136% average increase in confidence ratings between blocks 1 and 2

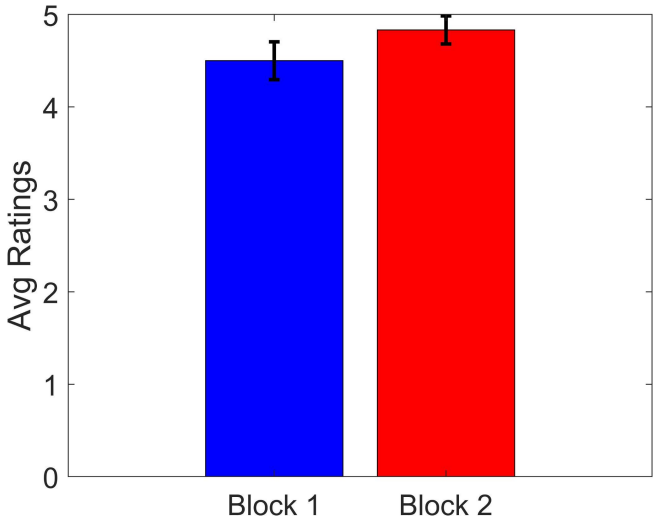

Fig. 4 Non-noticer group participants showed no significant increase in average confidence ratings between blocks 1 and 2.

False alarms (FAs) for noticer and non-noticer groups were analyzed separately in two 2 (block) x 3 (stimulus type) repeated-ANOVAs. For non-noticers, a significant decrease in the total number of false-alarms between blocks 1 and 2 was observed in the main effect of block $F(1, 23) = 18.47$, $p = < .001$) and stimulus type $F(2, 46) = 12.92$, $p = < .001$). In addition, the interaction between stimuli type and block was significant $F(2, 46) = 12.45$, $p = < .001$) (Fig. 5). A simple effect analysis across the means FAs for the 3 stimulus types in Block 1 was significant ($F(2,46) = 14.04$, $p < 0.0001$), and a Tukey's HSD showed that the mean false-alarms for SWS stimuli was significantly different from the mean FAs of the 2 other stimulus classes ($p < 0.01$) (no other means FAs were significantly different from each other). No significant differences in number of false-alarms across stimulus type or block were seen in the noticer group.

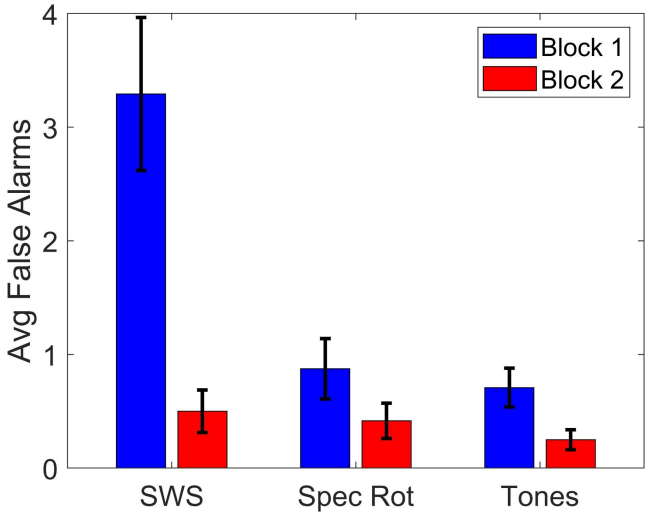

Fig. 5 Average difference in false alarm (FA) rates for non-noticers between blocks 1 and 2 show a significant decrease for SWS stimuli but not spectrally-rotated control words or pure tones.

Electrophysiological Results

Due to the variability in left-handed participant's hemispheric language lateralization, with right-hemispheric or bilateral language lateralization occurring more often in left-handed than right-handed subjects (Somers et al., 2015), we removed left-handed participants from final EEG analysis. A repeated measures ANOVA on the PAN voltages revealed a significant interaction between block and stimulus type for right-handed non-noticers $F(1, 23) = 4.45$, $p = .047$, $\eta^2_p = .175$ (Fig. 6). Figure 7 and 8 shows the grand-average ERPs averaged across all non-noticers for Blocks 1 (blue) and 2 (red) to spectrally rotated stimuli and SWS stimuli, respectively. Dotted vertical lines indicate the time window of the occurrence of the hypothesized "PAN" which is seen in Figure 8 as the slight negative shift in the Block 2 ERP during this period relative to the Block 1 ERP.

Scalp maps of differences in SWS ERPs between Blocks 1 and 2 showed a left-lateralized effect of hearing SWS as speech (Fig 9 left panel). Conversely, spectrally-rotated ERPs between blocks indicated no significant speech-related lateralization of negative voltage distribution (Fig 9 right panel).

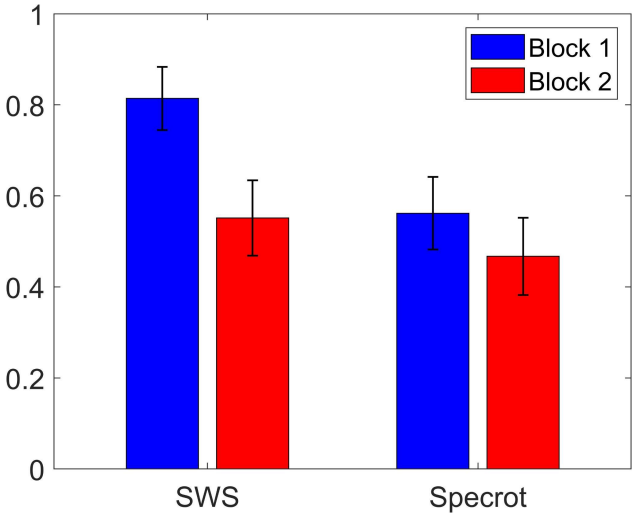

Fig. 6 Mean PAN averaged across all non-noticers for SWS compared to control stimuli between blocks 1 and 2 show a near-significant interaction between blocks 1 and 2 for SWS ERP averages.

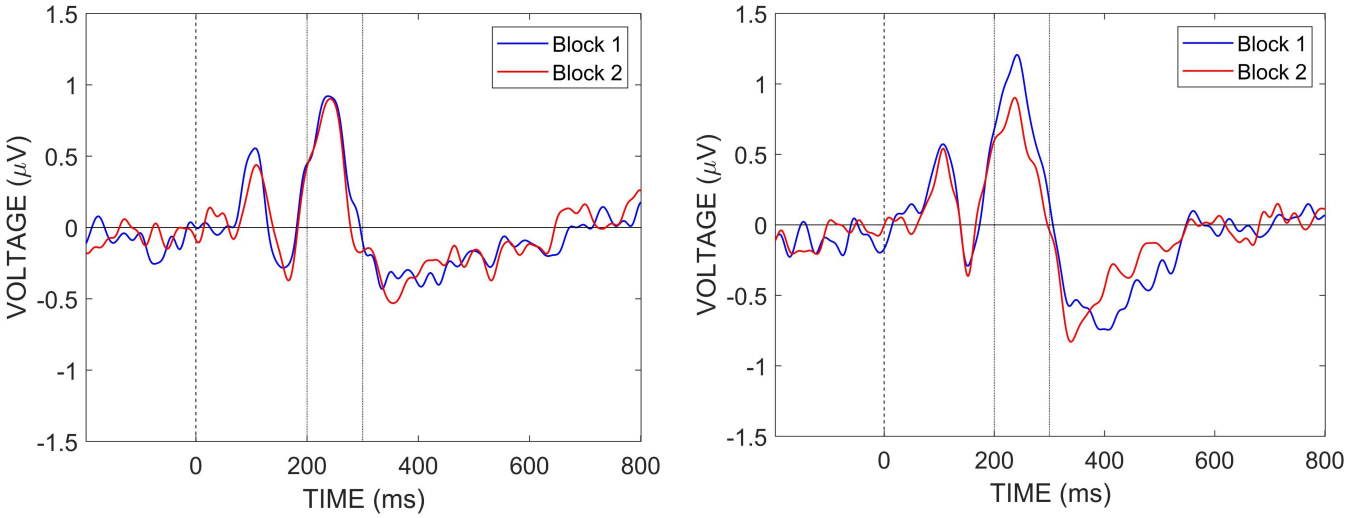

Fig. 7 200-300 ms post-stimulus time window for stimulus onset shows no significant changes in negative peak between blocks 1 and 2 for spectrally rotated control stimuli.

Fig. 8 200-300 ms post-stimulus time window shows a slight change in negative peak between blocks 1 and 2 for SWS stimuli.

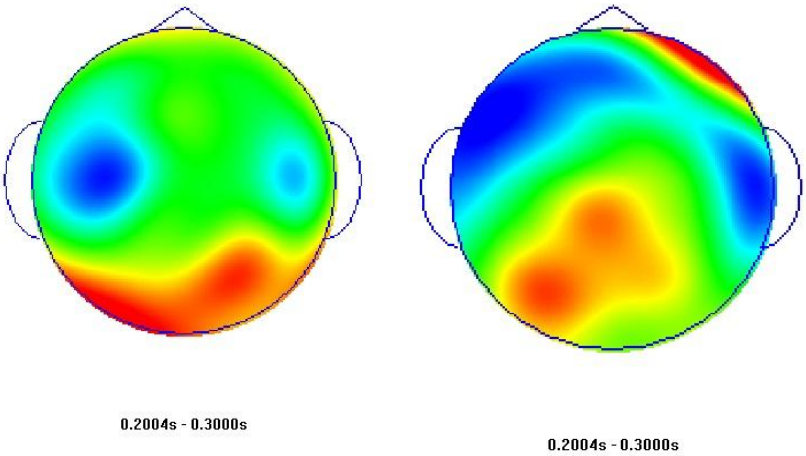

Fig. 9. Scalp map of differences in SWS (*left*) and spectrally rotated (SR) control word (*right*) ERPs between Blocks 1 and 2. The SWS map shows a left-lateralized effect of hearing SWS as speech, while the SR map does not.

Discussion

Summary of Results

Our behavioral results exhibited clear evidence of perceptual learning of SWS, with false-alarm rates decreasing across blocks for SWS exclusively, and average confidence ratings for hearing speech significantly increasing in non-noticers between blocks 1 and 2. After demonstrating that listeners indeed learned to hear speech in previously unidentifiable SWS stimuli, we analyzed the EEG data to see if there was also a corresponding neural signature of perceptual learning. To assess this, we sought to confirm Zhu et al. and Hendry's findings by analyzing ERPs across left fronto-central electrodes occuring in the 200-300 post-stimulus window after perceptual learning of SWS. We found a similar negative shift in the ERP in the 200-300 ms post-stimulus time window for non-noticers after speech training between Blocks 1 and 2, labeled by Zhu et al as the "perceptual awareness negativity" (PAN). At the same time, no similar negative shift was observed across blocks for these same listeners in the ERPs to spectrally-rotated control sounds, indicating the absence of a PAN for stimuli not perceived as speech. These results seem to successfully replicate the findings of Hendry (2019) and Zhu et al.

Theoretical Implications

The implications of the present study are contingent on the way the "perceptual awareness negativity" (PAN) is defined. ERPs showed a significant change in neural

activation after speech training for SWS, and scalp maps showed a language-related left-lateralized effect after learning to hear SWS as speech. Paired with strong behavioral evidence of enhancements to language intelligibility, there is an argument to be made that the PAN is a novel ERP component capable of measuring changes in speech (or perhaps perception, generally) awareness. However, several other ERP components share characteristics with the PAN and may offer alternative explanations.

The auditory mismatch negativity (MMN) is an evoked-potential identified as a negative deflection in voltage after presentation of an aberrant signal disrupts a regular pattern of repeated acoustic stimuli (Garrido, 2009). It has, like the PAN, a frontocentral localization. It may be the case that once subjects begin to recognize phonological elements in SWS stimuli that these perceptually-shifted events have started to act as deviant stimuli, causing a response like that of the MMN. Another component, the N200, has several sub-components which could partially explain the behavior of the PAN and lean towards a non-linguistic identity of the ERP. N2b has the same frontocentral location as the MMN, and roughly the same post-stimulus time window as the PAN. It is usually related with the P300 component, and is resultantly dependent on attention to target stimuli (Patel & Azzam, 2005). In the case of the present study, N2b can function as an error-monitoring ERP, and may indicate when a shift in attention occurs in response to a deviant stimulus such as when SWS emerges as a different kind of sound (subjectively) after speech-training.

These alternative explanations implicate the PAN in measuring attentional or perceptual changes in response to novel auditory stimuli. It is evident that perceptual learning of SWS is occurring in the study, but whether the PAN component specifically measures changes in perceiving SWS as intelligible speech is unclear. Future research using simultaneous recording techniques with improved spatial resolution along with changes to the existing design may provide clearer answers.

Limitations

The weaker PAN effect we observed in our data (even with a larger sample size) may be due to changes in the present design relative to Hendry and Zhu's method. We did not, for example, attempt to focus attention away from the SWS stimuli by using the pure-tone stimuli as targets in a one-back task. Possibly the biggest difference is that even though we had more subjects in our non-noticer group, we might have had fewer overall trials in the average ERPs because we used 2 stimuli of each stimulus type, rather than 3, and repeated each stimulus type only 50 times rather than 100. Therefore, we have less than half of the number of epochs forming our ERPs as Zhu et al., (*under review*) and Hendry (2019) do. This will make for noisier ERPs and hence could obscure the presence of a PAN across individual subjects. In addition, questionnaires in our study asked participants if they heard any speech content in the sounds they heard, and left out alternative sources for participants to choose from, such as 'environmental' or 'animal' sounds. Therefore, it's possible that listeners' experience of SWS stimuli was enhanced due to the content in questionnaires. Other unforeseen influences may have played a role in participant speech comprehension, such as the possibility that participants overhearing researchers talking outside of the sound booth during the recording session, acting as a form of speech priming. Lastly, spectrally rotated versions of 'Basic' were unsuccessfully rotated around a center frequency,

meaning that there was an imbalance between control words for ‘Basic’ and ‘Short’, which likely lowered the internal validity of the experiment. However, given that no subjects reported hearing a word for the present version of the spectrally rotated "basic", we are confident that it acted as a control comparison to a full spectrally-rotated version of the word.

Acknowledgements

I would like to thank Quinn Foti for his diligence, effort, insightful comments, and hands-on support throughout EEG data collection; Quinn’s assistance in capping participants and monitoring active EEG recording was extremely helpful, and his company alone was equally valued. I would also like to thank the UNCA Undergraduate Research program for funding this study and to all participants for their curiosity and willingness to sit through this odd and fun experiment. Finally, I would like to thank Dr. Michael Neelon for his academic and intellectual mentorship, and his constant encouragement and support. I am thankful for his teaching to me various skills, many of which were tangibly part of making this project possible, and for the countless intangible skills which kept me motivated to the end.

References

Beer, A. L., Batson, M. A., & Watanabe, T. (2011). Multisensory perceptual learning reshapes both fast and slow mechanisms of crossmodal processing. *Cognitive, Affective, & Behavioral Neuroscience, 11*(1), 1-12. 10.3758/s13415-010-0006-x

Bosker, H. R. (2018). Putting Laurel and Yanny in context. *The Journal of the Acoustical Society of America, 144*(6), EL503-EL508. https://doi.org/10.1121/1.5070144

Cohen, Jonathan, 'Perceptual Constancy', in Mohan Matthen (ed.), *The Oxford Handbook of Philosophy of Perception* (2015; online edn, Oxford Academic, 10 Sept. 2015), https://doi.org/10.1093/oxfordhb/9780199600472.013.014, accessed 14 Dec. 2023.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clinical neurophysiology, 120*(3), 453-463. https://doi.org/10.1016/j.clinph.2008.11.029

Hendry, C. (2019). Thesis for Bachelor of Arts, Reed College. https://www.reed.edu/psychology/scalp/thesis/files/Hendry_Thesis.pdf

Lake, B., Salakhutdinov, R., Gross, J., & Tenenbaum, J. (2011). One shot learning of simple visual concepts. *Proceedings of the annual meeting of the cognitive science society*, *33*(33).

Leibold, L. J. (2017). Speech perception in complex acoustic environments: Developmental effects. *Journal of Speech, Language, and Hearing Research, 60*(10), 3001-3008. https://doi.org/10.1044/2017_JSLHR-H-17-0070

Liebenthal, E., Binder, J. R., Piorkowski, R. L., & Remez, R. E. (2001). Sinewave speech/nonspeech perception: An fMRI study. *The Journal of the Acoustical Society of America, 109*(5), 2312–2313. [https://doi.org/10.1121/1https://doi.org/10.1121/1.4744123)

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

Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: selected studies of the event-related potential. *International journal of medical sciences, 2*(4), 147. https://doi.org/10.7150%2Fijms.2.147

Rodríguez, G., & Angulo, R. (2014). Simultaneous stimulus preexposure enhances human tactile perceptual learning. *Psicológica, 35*(1), 139-148.

Rubin, N., Nakayama, K., & Shapley, R. (2002). The role of insight in perceptual learning: Evidence from illusory contour perception. *Perceptual learning*, 235-251.

Saija, J. D., Akyürek, E. G., Andringa, T. C., & Başkent, D. (2014). Perceptual restoration of degraded speech is preserved with advancing age. *Journal of the Association for Research in Otolaryngology, 15*, 139-148. https://doi.org/10.1007%2Fs10162-013-0422-z

Sánchez, J., Dwyer, D. M., Honey, R. C., & de Brugada, I. (2022). Perceptual learning after rapidly alternating exposure to taste compounds: Assessment with different indices of generalization. *Journal of Experimental Psychology: Animal Learning and Cognition, 48*(3), 169. https://psycnet.apa.org/doi/10.1037/xan0000333

Seitz, A. R. (2017). Perceptual learning. *Current Biology, 27*(13), R631-R636. https://doi.org/10.1016/j.cub.2017.05.053

Sekar, K., Findley, W. M., & Llinás, R. R. (2012). Evidence for an all-or-none perceptual response: Single-trial analyses of magnetoencephalography signals indicate an abrupt transition between visual perception and its absence. *Neuroscience, 206*, 167-182. https://doi.org/10.1016/j.neuroscience.2011.09.060

Sohoglu, E., & Davis, M. H. (2020). Rapid computations of spectrotemporal prediction error support perception of degraded speech. *eLife, 9*, e58077. https://doi.org/10.7554/eLife.58077

Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42(4)*, 1096-1104. (**PDF**) https://doi.org/10.3758/brm.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24-31. https://doi.org/10.1177/0098628316677643

Wang, X., & Xu, L. (2021). Speech perception in noise: Masking and unmasking. *Journal of Otology, 16*(2), 109-119. https://doi.org/10.1016/j.joto.2020.12.001

Wong, P. C., Uppunda, A. K., Parrish, T. B., & Dhar, S. (2008). Cortical mechanisms of speech perception in noise. https://doi.org/10.1044/1092-4388(2008/075)

Zhu, Y., Li, Charlotte., Hendry, Camille., Glass, James., Canseco-Gonzalez, Enriqueta., Pitts, A. Michael., Dykstra, R. Andrew., (*under review*). Isolating neural

signatures of conscious speech perception with a no-report sine-wave speech paradigm

Appendix

Appendix A. Word selection for Pilot Study and EEG Experiment including Rudell (1993) ratings.

Word Selection	Syllable Count	Kucera-Francis (1967) Word Frequency	Mean Difficulty Rating	SD Difficulty Score
Above	2	296	-102	37
After*	2	1070	-104	35
Among	2	370	-32	39
Basic*	2	171	-14	49
Being*	2	712	-42	55
Black	1	203	-123	21
Blood	1	121	-109	31
Brief	1	73	-22	47
Bring	1	158	-109	32
Broad	1	84	-47	48
Cause	1	130	-38	67
Civil	2	91	54	46
Color	2	141	-123	21
Event	2	81	-18	50
Exist	2	59	-5	49
Faith	1	111	-2	40
Floor	1	158	-118	23
Force	1	230	-40	42
Front	1	221	-105	34

Given	2	377	-46	47
Green	1	116	-113	39
Hence*	1	58	62	43
Honor	2	66	-1	39
Horse	1	117	-123	21
Image	2	119	7	41
Index	2	81	37	50
Leave	1	205	-107	38
Legal	2	72	-18	45
Level	2	213	-14	32
Lived	1	115	-49	54
Meant	1	100	-22	67
Money	2	265	-101	34
Moral	2	142	19	42
Moved	1	181	-100	38
Music*	2	216	-118	23
Novel	2	59	16	49
Offer	2	80	-28	38
Paper	2	157	-123	21
Party	2	216	-102	37
Plant	1	125	-101	38

Range	1	160	-2	41
River	2	165	-106	34
Scene	1	106	-27	36
Serve	1	107	-50	40
Seven	2	113	-114	30
Short*	1	212	-102	41
Small	1	542	-115	31
Speak	1	110	-111	36
Speed	1	83	-42	45
Staff	1	113	2	43
Start	1	154	-106	35
Stock	1	147	8	41
Story	2	153	-116	37
Table	2	198	-123	21
Teeth	1	103	-118	26
Third	1	190	-100	37
Under	2	707	-111	35
Water	2	442	-118	26
Woman	2	224	-105	35
Wrong	1	129	-114	26