



FALSE HEARING AND THE N400: THE EFFECTS OF LINGUISTIC CONTEXT ON LANGUAGE PERCEPTION

**Mariah Erickson (Faculty Mentor: Brennan R. Payne PhD)
Department of Psychology**

ABSTRACT

False hearing is a phenomenon where one mishears what has been said to them based on linguistic contextual cues used to make a prediction (Rogers et al., 2012). The incorrect hearing usually has similar phonemic properties to other likely words and syntactic relation to what was said prior. Our study used methods from audiology and electrophysiology to analyze how linguistic contextual predictions impact perception in hearing. We observed the N400 response (Kutas & Hillyard, 1980) to incongruent phonologic lures used as target words that were presented in noise at a + 3dB SNR. We used the phonologic lure (PL) condition to collect data in instances where false hearing occurred (FH+) to compare to non-false hearing trials (FH-) as well as trials that used congruent (CON) or incongruent baseline (IB) words. We found a larger N400 effect in the IB condition compared to CON, and an intermediate N400 to the PL condition. Although our findings were not statistically significant, we did observe a trend of a decrease in amplitude in FH+ trials when contrasted to FH- trials. This may indicate that in false hearing, the participants response is somewhere in between hearing the congruent and the incongruent word with a very small threshold for choosing the word with a strong semantic fit to the perceived cue word over the actual incongruent phonologically similar target word. This suggests that false hearing may have a perceptual rather than a post-perceptual locus.

INTRODUCTION

Listeners can use linguistic context to facilitate word recognition (Bilger et al., 1984). One way they may do this is by using context to predict upcoming words. The process of making predictions is a top-down process, meaning that one is relying on prior understanding to interpret sensory perceptions. This is in contrast to a bottom-up process, which involves retrieving sensory information from one's environment to build on perception, e.g., hearing (Kutas & Federmeier, 2011). In the current study, we examined what happens when people misinterpret what is in their sensory environment based on top-down context-driven biases in spoken language. We wanted to determine if false hearing is due to either bottom-up processing of the stimulus being changed by an expectation; or, if it is due to context-based guessing at responding, i.e., not a result of perception but of being 'captured' by the context (Failes et al., 2020).

Our default state is to interpret as much potentially relevant information as possible with as little cognitive resources as possible. Because neural networks are already activated upon hearing one word (Bilger et al., 1984), filling in phonemic properties for upcoming words using a top-down process instead of a bottom-up process is highly efficient. However, reliance on context driven linguistic prediction may lead to greater instances of mishearing when the

conditions are right. This mishearing phenomenon is known as false hearing (Rogers et al., 2012), or instances where the misperceived word has some rational semantic relation to the available context. For instance, if someone were to say, “my nephew kicked my knees.” As part of recognizing the word “nephew” we activate lexical features of words that are semantically associated with the word nephew. This could be things semantically related, like other terms associated with family, and in turn the brain prepares to hear these associated words by activating the phonemic properties for them (Elman & McClelland, 1983). This might lead the listener to mishear, e.g., instead of hearing “knees” they hear “niece.” When context is being used to form word level predictions the occurrence of mishearing is due to a stronger semantic association (i.e., semantic relation to “nephew”: niece > knees) between a target word and a word that sounds similar to what was really said (i.e., a phonologic relation: niece (International Phonetic Alphabet (IPA): ni:s) and knees (IPA: ni:z) share similar phonological onsets) (DEESE, 1959; Nelson et al., 2004).

False hearing has been studied using behavioral measures by Rogers et al. (Rogers, 2017; Rogers et al., 2012; Rogers & Wingfield, 2015). The study we based our experimental design off of, Rogers et al. (2012), used a noise masking technique where the level of background noise is manipulated during presentation of the target, i.e., less or more noise depending on the condition. Participants responded to a trial by stating what they had heard and indicated how confident they were that they had identified the word correctly. The accuracy of hearing and levels of noise used to mask target words showed to have a strong negative correlation. As one would expect, a higher signal to noise ratio (SNR) resulted in lower accuracy in perception and recall.

Using sound masking techniques to distort an acoustic speech-signal is effective for eliciting greater effort in language perception (Pichora-Fuller et al., 2016; Silcox & Payne, 2021; Tun & Wingfield, 1999). The quality of an acoustic speech-signal is fundamental for accuracy of hearing (Rogers et al., 2012). It is likely one has experienced that there is a significant difference between areas that are and are not designated for effective communication. Having a quiet acoustic environment helps speech perception. For example, auditoriums, lecture halls, and classrooms are some places where hearing with ease is one of the essential features architects keep in mind in the design process. Clarity of the sound signal can be distorted by interfering noise, leading to difficulty hearing. Poor sound quality compromises hearing which can lead to difficulty recognizing spoken words, remembering the details of the conversation, or, if the conditions are right, perceiving the wrong word entirely, i.e., false hearing (Rogers et al., 2012).

One of the consequences of a noisy acoustic environment is that it increases the effort needed to listen and interpret sound-signals (Rabbitt, 1968). In circumstances where greater effort is needed to interpret speech, like when exposed to background noise that distorts a sound signal, this is referred to as effortful listening. Our understanding of effortful listening is taken from a framework called FUEL or Framework for Understanding Effortful Listening, proposed in the fifth Eriksholm Workshop on *Hearing Impairment and Cognitive Energy* (Pichora-Fuller et al., 2016), based off of Kahneman’s original Capacity Model of Attention (Kahneman, 1973; Pichora-Fuller et al., 2016). FUEL proposes that each individual has a limited capacity of cognitive resources for completing a task, and that the threshold of resources is varied from person to person. How demanding a task is corresponds to the amount of cognitive resources used to complete it. In effortful listening conditions, the resources used to successfully perceive speech may be taken away from other higher-level processes like encoding what is heard into memory (Payne et al., 2019; Rabbitt, 1968; Silcox & Payne, 2021).

Evidence found to support the theory that effortful listening decreases memory of what was said was found in a classic study by Patrick M. Rabbitt (1968). In the Rabbitt study participants did a speech perception is noise-task and were asked to remember two lists of numbers presented either with background noise or in quiet. There were three groups: group A: quiet, quiet; group B: quiet, noise; and group C: noise, noise. What they found was that exposure to the noise condition interfered with retaining earlier information that was not acoustically challenging. This was determined due to later recall of the first list in group B (quiet, noise) being significantly worse than in group A (quiet, quiet) and as bad as group C (noise, noise). This finding is striking as the first list in Groups A and B were both presented in quiet, thus, hearing the second list in noise in Group B impacted recall of the first list. This demonstrates the negative effects of effortful listening. The findings from the Rabbitt (1968) study show how effortful listening effects more than just accuracy in hearing.

In McCoy et al. (2005), Rabbitt's (1968) findings were extended by using a running memory span task on a group of older adults with normal hearing and a group of older adults with poor hearing. The task involved listening to a continuous list of words while being periodically interrupted and asked to repeat back the last three words that were heard. Both the hearing-impaired and non-hearing-impaired groups were at or near ceiling levels of accuracy in their recollection of the most recent words, indicating that both groups were correctly perceiving each word in the list. However, they found that the older adults with normal hearing performed significantly better than those with hearing impairment at correctly identifying the first two words. McCoy et al. (2005) concluded that the discrepancy in later word recall supports the hypothesis that effortful listening comes at the cost of worse maintenance of information in verbal memory.

Although poor acoustic conditions are problematic, the brain has a way of alleviating the negative effects sound interference has on accuracy and memory. As was seen in Rabbitt (1968), added effort for hearing pulls resources away from other higher-level cognitive processes. Thus, the older adults in McCoy et al. (2005) with poorer hearing were accurately perceiving the words but not encoding them into memory as efficiently (Payne et al., 2019; Pichora-Fuller et al., 2016; Tun & Wingfield, 1999). Importantly, McCoy et al. (2005) found that the negative effects of poor listening conditions on memory could be offset by the availability of linguistic contextual cues. They showed that when the word sets included congruent words, recall of all three words was at or near ceiling. They also found that in the group with hearing loss, recall of the final word was even higher than in the group with normal hearing. This may be due to those with hearing impairment having to rely on context more frequently, leading to a heightened ability for forming linguistic predictions (McCoy et al., 2005).

In another study contextual cues were used in a revised speech perception in noise test, known as the R-SPIN, to test the effect of context on memory (Gordon-Salant & Fitzgibbons, 1997). The R-SPIN test includes 200 low predictability and high predictability sentences. As an example of high and low predictability sentences, given the phrase "I take my coffee with cream and....," most people would fill in the final sentence word with "sugar". Because there is a limited amount of final sentence words to the above example, this is a high predictability sentence, whereas "All day long she thought about..." is a low predictability sentence because the final word could be anything. In the study (Gordon-Salant & Fitzgibbons, 1997), the participant groups were composed of four different categories: young adults with normal hearing, older adults with normal hearing, young adults with hearing impairment, and older adults with hearing impairment. They found that the groups with hearing impairment performed

significantly worse in all trials excluding trials where contextual cues were available. However, in high predictability trials participants in both groups had nearly perfect accuracy in word recall and in sentence recall.

There is a large body of supporting evidence on how linguistic context helps to improve language perception, with more contextual cues leading to less cognitive effort needed for listening. However, linguistic context may also lead to a greater occurrence of false hearing (Rogers et al., 2012; Rogers & Wingfield, 2015; Silcox & Payne, 2021). The Rogers and colleagues false hearing studies often separate groups by younger adults and older adults. Interestingly, the groups of older adults consistently show to have higher accuracy in conditions with contextual cues, as was also seen in Gordon-Salant et al. (1997). This indicates that age has some relation to the extent context is relied upon for hearing, possibly due to declines in hearing or conditioning over time (McCoy et al., 2005). However, groups of older adults also showed a higher rate of false hearing (Rogers et al., 2012). The differences in age groups determines reliance on linguistic context, which is believed to increase with age, leading to increased susceptibility to false hearing. Even though false hearing showed to be more prevalent in groups of older adults, younger adults still showed to be susceptible to the same errors.

Another interesting study done by Failes et al. (2020), theorized that false hearing is due to a phenomenon referred to as the capture effect. The capture effect theorizes that the bottom-up process of hearing is less attended to due to being captured by the available linguistic contextual cues. This conclusion was in part due to past literature (Jacoby et al., 2005) where using predictive measures for capture, recollection, availability bias, and word generation provided a more accurate model for their data than without including measures for a capture effect (Failes et al., 2020; Jacoby et al., 2005). In Failes et al. experiment one, participants heard a sentence without noise and withheld the sentence final word as the target word played with background noise. Participants then would type the sentence final word they heard and select whether they remember hearing the word, know they heard it, or guessed. In experiment two, participants studied congruent semantically related word pairs and were told they would do a memory task afterwards on them. In the memory task, participants saw a prime word that was either the correct word (skull), an alternative word which could fill in the blank (scalp), or a string of five ampersands (&&&&&), followed by a cue–fragment displayed on a screen (e.g., head–s--l-). Participants then filled in their best guess as to what the actual word was and indicated if they remember that word specifically, did not remember specifically but were still sure it was the correct word, or guessed.

Failes et al. (2020) after correcting for differences in hearing between groups of younger and older adults found that accuracy in hearing and recollection in the baseline condition was better for older adults in both experiments than it was for younger adults. This would discredit the theory that increases in false hearing are due to declines in memory and hearing, but instead may be due to declines in cognitive control. In Balota and Spieler (1999), they theorize that this lack of inhibitory cognitive control is due to a prepotent response activating a spanning tree network from an associated prime rather than retrieval through recollection.

The N400 ERP Component

To observe in more detail how linguistic contextual predictions may lead to false hearing we used event related potentials to examine rapid neural responses during word perception. Using ERP's in the psychophysiological study of language processing has been a common and

reliable method of neural imaging in research for some time (Kutas & Hillyard, 1980). An ERP is composed of averaging recordings of brain electrical activity in high temporal resolution (milliseconds) via electroencephalography (EEG). These recordings are time locked to an event of interest (e.g., presentation of a stimulus, motor response), allowing for observation of the reaction to the stimuli. The N400, originally discovered by Kutas and Hillyard (1980), can be used to observe the semantic fit of a target word. The N400 is an ERP component which is named as such in reference to a negatively polarized electromagnetic deflection in neural activity, occurring around 200 - 600ms (milliseconds) from a target onset, which generally peaks around 400ms (Kutas et al., 1988). The N400 is of particular interest in our study due to its association with cognitive resources used to access semantic information. The larger the peak of the N400, the more cognitive resources activated to retrieve semantic information associated with the stimuli being processed. When a prediction involving possible upcoming words that would fit the context is made and what is heard confirms that prediction, we see a smaller absolute value of the N400. Therefore, we consistently see more semantically probable stimuli with a low amplitude N400, and unpredictable stimuli with a higher amplitude N400 (Kutas & Federmeier, 2011).

The association between the N400 and level of semantic fit is theorized due to the N400's sensitivity to the use of incongruent final sentence words. For example, the N400 to eyebrows is smaller in "He shaved his mustache and *eyebrows*" than the anomalous word *city* in "He shaved his mustache and *city*". However, the final word in "He shaved his mustache and BEARD." or "He shaved his mustache and beards.", do not elicit a differential N400 response (Kutas & Hillyard, 1980), suggesting that it is not due to general surprisal, i.e., visually or grammatically unexpected, but rather is driven by semantic expectation.

The Current Study

The N400 can be used to determine if what is observed was semantically expected or not, i.e., fulfilled a prediction made by the listener. We used this to our advantage to observe if words in the false hearing paradigm were being perceived by participants either as congruent or incongruent as a function of their self-reported perception. This should indicate, in instances of false hearing, if a target word was genuinely misheard by a participant, or if correct hearing is overridden in favor of a word with a better contextual fit at the time of responding (i.e., a *capture effect*; (Failes et al., 2020). To the best of our knowledge, using the N400 to observe participant responses to false hearing has not been previously examined. Therefore, the following study has implications of gaining valuable insight into the use of prediction in speech processing (Silcox & Payne, 2021).

If misperception of the target word in the PL condition elicits a smaller N400 response, meaning the participant misperceived the actual word for an expected word, this would give evidence that listeners use linguistic contextual predictions to ignore later phonemic sound signals of auditory speech once certain sound characteristics are processed that match their prediction. If participants have a larger N400 response to the target word in the PL condition, meaning participants did hear correctly, it may show that incongruity is perceived but perhaps does not make it through an attentional filter to recognition, or is over-ridden by a strong top-down expectation at the time of response. If we see misperception of the target word occurs after hearing correctly, it would support the possibility that the brain accounts for errors in hearing by ignoring information which is flagged as less likely (Batterink & Neville, 2013), or which cannot be semantically linked to any sensical phonemic pairings (Blackford et al., 2012).

MATERIAL & METHODS

Participants

The study included 15 participants, young adults (male: 6, female: 7, other: 1, mean age = 19). Participants were recruited from the University of Utah participant pool. The 15 participants were given class credits for their time participating. The screening for study eligibility confirmed the participants were predominantly right-handed (Oldfield, 1971), as well as had no past instances of traumatic brain injury, and reported English as their first language with no exposure to a second language before the age of 7. We assessed hearing acuity using pure tone audiometry and speech reception threshold tests in each ear via a modified Hughson-Westlake pure tone identification procedure. All participants had normal hearing. Controlling these participant characteristics reduces the likelihood for EEG differences that could potentially impact the interpretation of our study data.

Materials & Design

A list of 120 different word sets were used. Sets included one cue word and three paired target words which contain only one syllable. An example of some sets labeled with which word belongs to each target word condition is shown in Table 1. The conditions of target words were congruent (CON; e.g. cue: *half*, target: *whole*), incongruent baseline (IB; e.g. cue: *half*, target: *talk*), or they were incongruent but sounded similar to the congruent target word, which is referred to as “phonologic lure” (PL; e.g. cue: *half*, target: *home*). Participants heard 40 trials for each condition, 120 in total, with the cue–target pairing being counterbalanced across participants so that each cue word of the 120 was heard once by each participant. The presentation of conditions were randomized across trials, to avoid participants recognizing a pattern of condition delivery.

Table 1

Example of Study Stimulus Word Pairs

Cue Word	CON (Congruent)	PL (Phonologic Lure)	IB (Incongruent Baseline)
half	whole	home	talk
atlas	map	mat	lit
ill	sick	sin	food

Note. Example of the word pairs used in the experiment. Adapted from (Rogers, 2017). The first column is the cue word that the participants heard before a target word. Participants heard all cue words and one of the 3 target words for a total of 40 of each condition for each participant.

The audio stimuli were created using a male native speaker of American English and Adobe Audition software, with an audio sampling rate of 44.1 kHz (kilohertz). In MATLAB (MathWorks., 1990), a power spectrum matched noise masker set at + 3 dB (decibel) below the speech signal was added to all target word audio files. The signal to noise ratio (SNR) was to mask sound enough to increase listening effort while avoiding impairing intelligibility (Payne et

al., 2019). The chosen SNR is lower than what is used by Rogers et al. (Rogers, 2017; Rogers et al., 2012; Rogers & Wingfield, 2015) in their effortful listening tasks, but is challenging enough to increase effort needed for hearing and was chosen due to it being closer to the noise levels listeners would most likely experience outside of a controlled experiment. Moreover, prior work has shown that in young normal hearing listeners the chosen SNR increases listening effort without sacrificing intelligibility (Payne et al., 2021; Silcox & Payne, 2021).

The word pair list used in our experimental design was an adapted version of the word pair list used in Rogers (2017) which was created using a method known as forward association that assess semantic overlap between the given cue and target word. The forward association method was originally developed by Nelson et al. (2004) who gave participants a category or word, and then had the participants list off words that they related with the given prompt. The frequency of participant responses were then used to calculate the probability of an individual responding with one word given any other recorded word. This is much like what is seen on the popular game show *Family Feud*, where participants are asked to guess the most likely response that others would give to a prompt. For an example database of word forward association levels and more on the process of calculating forward association see the University of South Florida Free Association Norms (Nelson et al., 2004).

Procedure

After passing the initial screenings, described under the Participants section, the participants were administered several neuropsychological tests to assess their language proficiency. Including a short-form computerized version of the reading span task or RSPAN (Oswald et al., 2015). A modified FAS phonemic fluency test (Benton, 1968) which is a 60 second timed task that involves naming words that begin with a given letter, excluding proper nouns. And an extended range vocabulary test (Oswald et al., 2015; Payne et al., 2015).

Before beginning, the participants were instructed to turn off any electronic devices and store them in a separate room to avoid unnecessary distraction or electrical interference with the EEG recordings. The EEG cap was fitted to the participant, and they then completed a practice run to assure they understood the task. The trials took place in a separate room designed with the intention of limiting any outside distractions. Participants listened to the cue word without background noise followed by the target word embedded in background noise (SNR: +3 dB). The stimuli were presented diotically through insert earphones at a comfortable listening level of 65 dB hearing level. Participants were then asked to repeat the target word they had heard followed by a percentage to indicate how confident they were that they heard correctly. The average length of time to complete the study was approximately 2 hours and 20 minutes.

Participant responses were manually recorded by two separate lab researchers who were not exposed to the experimental materials to remove any potential bias in hearing. For any instances of differences in reporting participant responses another party was brought in to listen to participant responses for those trials to determine the participant response.

EEG Recording & Processing

We used a 32-channel, silver-silver chloride actiCap with slim active electrodes to collect EEG recordings from the participants (Brain Vision, LLC, Morrisville, NC, United States of America) in an international 10-20 montage. The processed EEG recordings for electrode Cz

were used, given that the N400 is known to be the most prevalent at this site (Kutas & Hillyard, 1980). Impedance levels were monitored, and recording stopped to adjust the cap if levels rose over 20 k Ω at any of the electrode sites. Reference electrodes were used (TP10 and TP9) near the left and right mastoids, i.e., behind both ears, TP10 for online reference and mean of TP10 and TP9 for offline reference. An electrode was placed under the left eye on the infraorbital ridge to create a channel for vertical eye movement (VEOG). TP10 was referenced to FT9 and FT10 to create a horizontal eye movement channel (HEOG). A BrainAmp DC amplifier was used for continuous EEG amplification with a cutoff at DC 0 Hz and an online low pass filter of 1000 Hz with a sampling rate of 500 Hz. EEG data were bandpass filtered offline at .1-30 Hz prior to analysis and were down sampled offline to 250 Hz.

In MATLAB/EEGLAB (MathWorks., 1990), the EEG data were processed for artifacts, e.g., blinking or eye movement that may affect results. Data were segmented at -100ms relative to target onset to 1200ms following target word onset. Artifact detection was done by using VEOG and HEOG, to detect shift, and flatlining. Thresholds were set, defaulting at 500 μ V (microvolts) for flatlining, 150 μ V for VEOG and HEOG, and 100 μ V for shift. Thresholds were then adjusted per participant until all significant artifacts were flagged and then removed in the final epoched data. In total, 22% of trials were rejected due to artifacts. No participants were excluded from the final results.

Data Analysis Pipeline

In R, the behavioral data, which consisted of all participant responses by the word sets (shown in Table 1) which were numbered from 1-120, were combined with the epoched EEG data, which were in chronological order, by reordering the behavioral data using a unique identifier for each trial.

The behavioral measures of interest were the average value for accuracy and confidence within conditions (i.e., CON, IB, and PL), while the EEG data contrasts used an a priori time window for the N400 of 250ms – 650ms from target onset. This is about 50ms later than the typical N400 time window due to a later onset of the N400 seen in auditory stimuli than in visually presented stimuli (Kutas et al., 1988).

Subject level averages ($n = 15$) for CON, IB, and PL were created for the reported confidence, average accuracy, and the N400 mean amplitude. Data were then analyzed using a repeated measures ANOVA for each outcome measure. Post-hoc paired t -tests were done following ANOVAs (seen in Tables 2 and 3).

For the false hearing-analyses, in the PL condition, trials were separated by FH+ (false hearing positive trial) and FH- (false hearing negative trial). Pairwise comparisons were used to contrast accuracy in CON, IB, and FH- trials, excluding FH+ where all trial accuracy was 0%. As well as for confidence in CON and IB to FH- and FH+ trials, reported under false hearing in the results section.

RESULTS

Results from Behavioral Data

The repeated measures ANOVA revealed a significant omnibus effect of condition on accuracy, $F(2, 28) = 63$, $p < 0.001$. Follow-up pairwise contrasts, as shown in Table 2, found

that accuracy was significantly higher in CON ($M = 91\%$, $SD = 28\%$) compared to PL ($M = 61\%$, $SD = 49\%$) and IB ($M = 70\%$, $SD = 36\%$). Note that accuracy was higher for IB trials than PL trials, meaning that participants were more accurate with their responses to incongruent words when they had no phonologic similarities to the congruent word, suggesting false hearing may be driving lower accuracy in the PL condition.

A repeated measures ANOVA revealed a significant omnibus effect of condition for reported confidence, $F(2, 28) = 30.6$, $p < 0.001$. Post-hoc analysis, seen in Table 2, found that the highest reported confidence was for CON ($M = 90.6\%$, $SD = 24\%$), followed by PL ($M = 75.3\%$, $SD = 33.5\%$), and then IB ($M = 69\%$, $SD = 46\%$). Note that although accuracy was lower for PL compared to IB, confidence for PL was higher than IB, consistent with false hearing (Rogers et al., 2012).

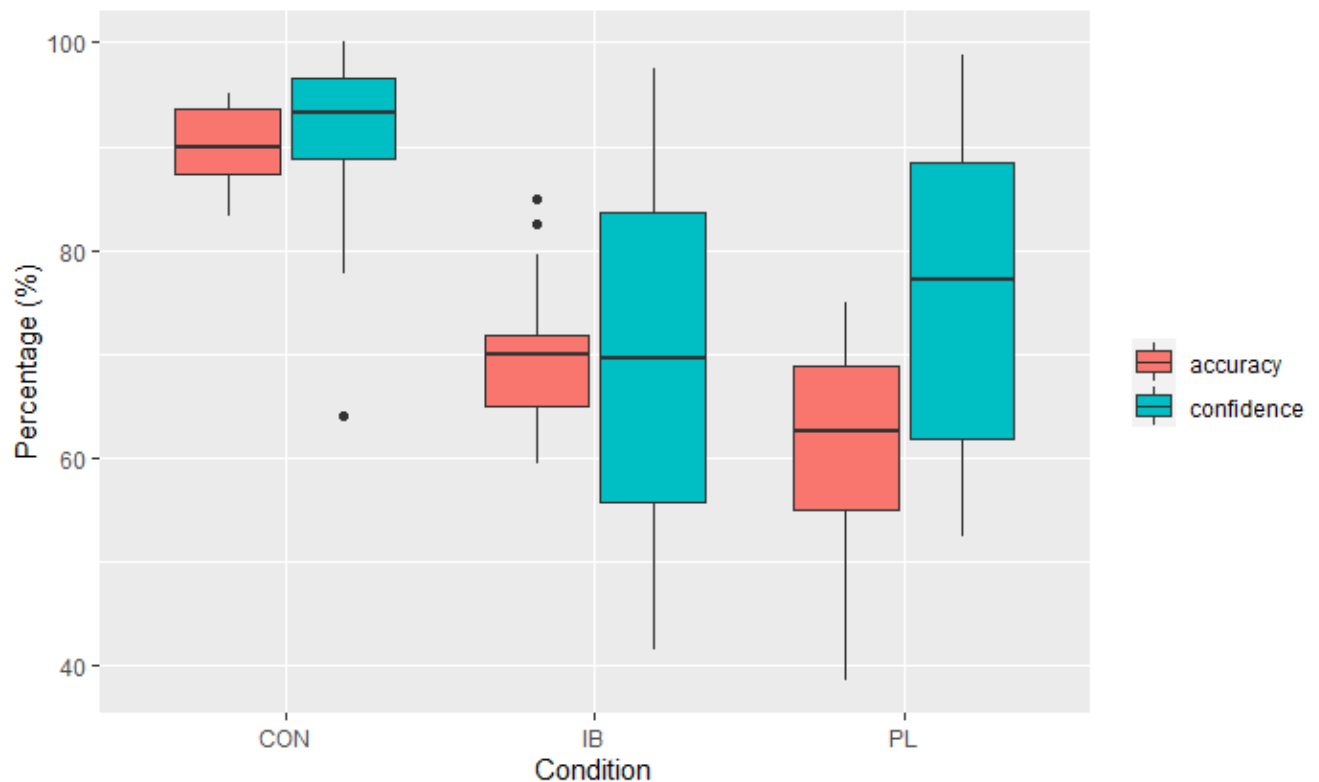
Table 2
Pairwise Contrasts for Behavioral Results

Contrast	$t(df = 14)$	p -value	Est. Dif.	95% CI
Accuracy				
CON versus PL	10.8	< 0.001	29.2	[23.4, 35]
CON versus IB	8.1	< 0.001	20.2	[14.8, 25.55]
PL versus IB	-3.25	< 0.01	-9	[-14.94, -3.05]
Confidence				
CON versus PL	5.8	< 0.001	15.06	[9.5, 20.64]
CON versus IB	5.9	< 0.001	20.83	[13.21, 28.45]
PL versus IB	3.2	< 0.01	5.8	[1.9, 9.7]

Note. Post-hoc contrasts of behavioral data for the conditions used: congruent (CON), incongruent baseline (IB), and phonologic lure (PL). The first table section, accuracy, shows pairwise contrasts for the percentage of correct responses given by averaging correct responses by participant and within the conditions. Confidence, in the second section, shows pairwise comparisons which were given as a percentage reported by the participant for each trial based on if they think they gave the correct response (100%), were sure they heard incorrectly (0%), or were completely uncertain (50%).

Figure 1

Accuracy and Confidence in Word Recognition as a Function of Condition



Note. Side by side visual representation of the differences seen between results for accuracy and confidence. Accuracy is seen on the left, in red, and confidence on the right, in blue. The X axis is grouped by the experimental conditions, congruent, incongruent baseline, and phonologic lure. The Y axis shows percent accuracy or confidence. Note the reverse relation of accuracy and confidence in the incongruent baseline and the phonologic lure conditions. Data within one *SD* (standard deviation of the sample) is indicated by the colored boxes for accuracy and confidence. Within two *SD* are indicated by the error bars. The outlier data points, indicated by the points seen in confidence under CON and accuracy in IB, are within three *SD*.

Results from EEG Data

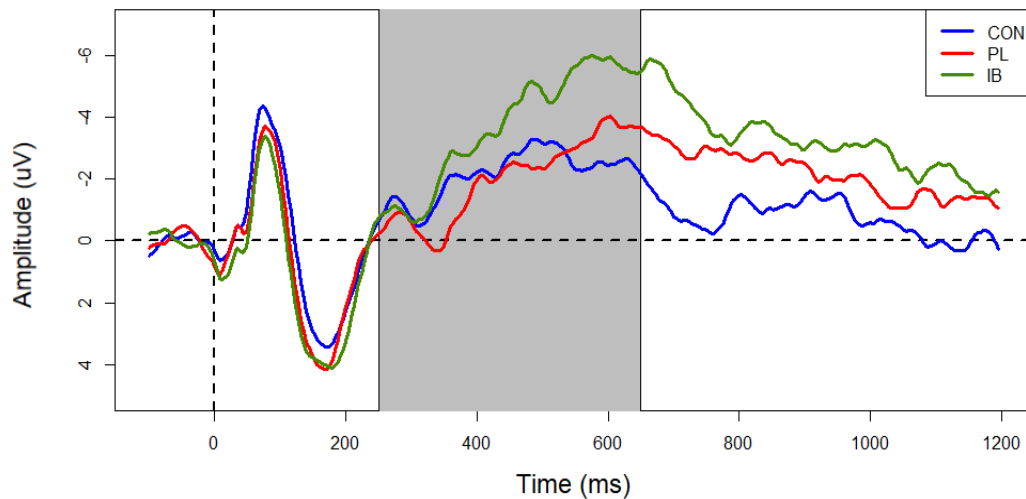
A repeated measures ANOVA revealed a significant omnibus effect of condition on the mean amplitude of the N400; $F(2, 28) = 3.5, p < 0.05$. The lowest (i.e., most positive) N400 mean amplitude was seen in CON ($M = -2.16, SD = 0.77$), followed by PL ($M = -2.02, SD = 1.31$), and highest mean amplitude in IB ($M = -3.73, SD = 1.84$). Table 3 shows the pairwise contrasts between conditions. IB was significantly different from both PL and, as expected, CON. However, the PL condition was not statistically different from CON.

Table 3
N400 Mean Amplitude Experimental Data Contrasts

Contrast	$t(df = 14)$	p -value	Est. Dif.	95% CI
CON versus PL	-0.07	> 0.05	-0.05	[-1.5, 1.4]
CON versus IB	2.17	< 0.05	1.56	[0.01, 3.12]
PL versus IB	2.25	< 0.05	1.61	[0.07, 3.16]

Note. The time window used to analyze the N400 mean amplitude was from 250 - 650ms. This table lists the pairwise post-hoc contrasts. For each contrast, the reported t -score is given, and the degrees of freedom used was 14 ($n = 15$). Values are reported as being less than $\alpha = 0.05$, 0.01, and 0.001. The average of the congruent and phonologic lure condition, shown in the first row, were most similar; IB was significantly higher in contrast to both the congruent and phonologic lure condition.

Figure 2
ERP-Cz: By Primary Experimental Conditions.



Note. The X axis shows the passage of time in milliseconds (ms), while the Y axis shows amplitude at each point of time in microvolts (μV). The time window used to calculate the N400 mean amplitude (250 – 650ms) is shown in grey.

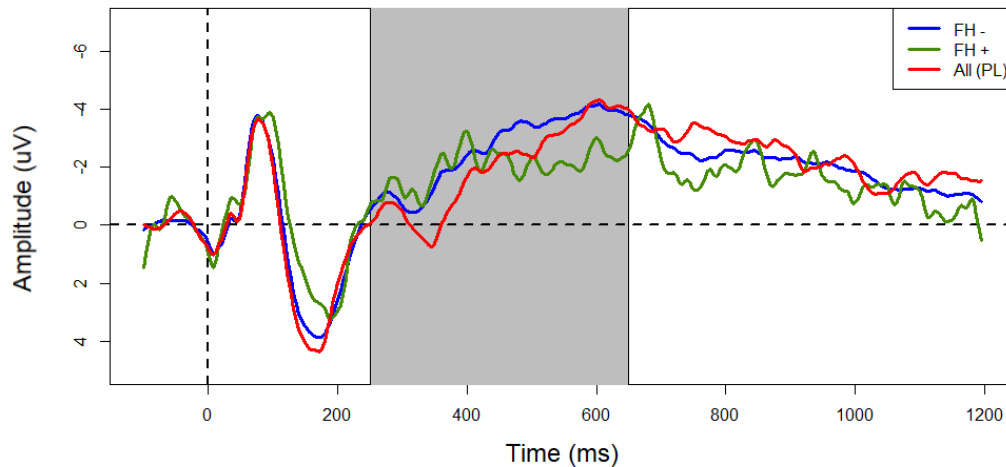
Results from False Hearing Data

To examine false hearing, the PL condition was separated into FH+ trials which were determined by the participant response being the same as the congruent target word from that word set, and FH- trials which include any other response given. A total of 109 trials were flagged as FH+ (false hearing rate = 18%), out of 600 total PL trials collected. Of the 109 FH+ trials, 26 were rejected due to artifacts. Final analysis on FH+ included 83 trials, an average of 5

trials per a participant. For FH+, mean reported confidence was higher ($M = 87.7\%$, $SD = 24.8\%$), compared to FH- trials ($M = 72.4\%$, $SD = 34.55\%$), but the difference was not statistically significant, $t(14) = 1.34$, $p = 0.25$. Accuracy for FH- was ($M = 75\%$, $SD = 43\%$). Figure 3 shows the ERP waveforms during target word processing separately for subsequently categorized FH+ and FH- trials. There was a trend for a smaller N400 amplitude to FH+ trials than FH- trials. However, contrasts for FH+ and FH- for the mean amplitude of the N400 were not statistically significant $t(14) = -0.51$, $p = 0.6$, estimated difference = -0.47 , 95% CI $[-2.45, 1.5]$.

Figure 3

ERP-Cz: By False Hearing Positive and Negative Trials.



Note. The X axis plots the passage of time in milliseconds (ms), while the Y axis plots amplitude in microvolts (μV). The time window used to calculate the N400 mean amplitude (250 – 650ms) is shown in grey.

DISCUSSION

Our aim was to determine how linguistic contextual predictions effect occurrences of false hearing. We hypothesized that false hearing was either due to a perceptual locus (i.e., false perception during word processing) or is due to a post-perceptual locus (i.e., context-based guessing after the PL target word was heard correctly but the phonemic characteristics that do not fit are dismissed (Failes et al., 2020). In the first scenario, top-down expectations hinder bottom-up perception based on the fit between an expectation and partially overlapping acoustic information. In the second scenario, an incongruent word is perceived early during word recognition, but this is overridden during response.

We sought to determine this through analyzing the N400 response to target words presented in noise. Unfortunately, our results were inconclusive, as to be expected with a preliminary sample size. Our data also showed that the N400 tended to be lower in FH+ trials than in FH- trials, which if this pattern is seen after more data is collected it would provide evidence that perception in the PL condition was most similar to the perception of the congruent word. After a larger sample has been collected, we will run comparisons on FH+ and FH- trials again. With further analyses we may learn more about how a decision is made in the PL condition which determines hearing as FH+ or FH-. With precise enough data, we may even be

able to determine a threshold for false hearing which leads to accepting the congruent word over the incongruent phonologically similar word.

In our results from analyses on our behavioral data, we observed a higher level of accuracy and confidence in CON trials, where the target word was congruent with the cue word. Near ceiling accuracy was also observed in Gordon-Salant and Fitzgibbons (1997) on trials that provided contextual cues, for both hearing impaired and non-hearing impaired participants. Our conclusion from results in the CON condition and findings of low accuracy in the IB condition is to agree with Gordon-Salant and Fitzgibbons findings that an increase in available contextual cues helps to significantly alleviate the negative effects of effortful listening on language perception. The increased accuracy in CON trials also supports theories of pre-activation (Batterink & Neville, 2013) which posits that upon hearing a single word, other semantically associated words are activated, resulting in faster recognition with a lower cognitive load than when relying on an acoustic speech signal alone. Pre-activation also lends to the theory that word recognition is hindered when a target word is incongruent with the context but shares phonologic similarities to a word with a strong semantic relation with the context (Silcox & Payne, 2021).

Although it should be uncontested that linguistic context improves speech perception, the negative effects of a reliance on contextual cues is an important part of understanding how context based linguistic predictions are applied in language perception. In our results, we saw an example of these negative effects in PL which had lower accuracy and higher confidence than was seen in IB. This is likely due to the similarities in the phonemic onset of the target word in PL and CON, which resulted in false hearing occurring in approximately one fifth of PL trials.

As was proposed by Balota et al. (1999) a lack of inhibitory control may lead to instances of falsely recalling words or events due to spreading activation or the pre-activation of semantically related information. The negative relationship between confidence and accuracy in the PL and IB condition illustrates the possibility that false hearing is caused by an inability to stop filtering of bottom-up perception of hearing from happening due to the covert nature of language filtering, also discussed as the capture effect in Failes et al. (2020).

In our analyses of the EEG data, our results indicate that listeners pre-activated phonological features of more likely upcoming words, consistent with past literature on false hearing (Rogers, 2017; Rogers et al., 2012; Rogers & Wingfield, 2015). Evidence for this is seen in the graded response in the N400 which was highest in the IB condition and lowest in CON. Evidence which relate strength of electrophysiological activity to cognitive effort, and the correlation of electrophysiological activity categorized as the N400 to predicting semantic probability seen in Kutas and Hillyard (1980) would lead to the logical conclusion that the lower amplitude in the congruent condition indicates that retrieval of congruent information takes less cognitive effort. Meaning, once participants were exposed to the cue word, rapid pre-activation occurred of a spanning tree network with associated concepts which includes phonemic properties of the most likely words. Participants are then captured by this pre-activation and unless hearing is consciously attenuated in the PL condition, this leads to falsely perceiving the target with higher confidence.

After comparisons of FH+ trials and FH- trials within the PL condition, we did not find any significant difference between conditions. However, listeners seemed to realize on some level that the phonologic lure was not congruent, indicated by the deviation seen later in the N400 waveform between CON and PL (shown in figure 2). Although the waveform differences

in the N400 for CON and PL did not prove to be statistically significant, they showed a distinctly independent wave form from each other. Future work should focus on more temporally specific analyses to statistically characterize the time-course of these effects mapped with the duration of the target word so we can see at what point exactly the word is recognized. For instance, is the later deflection seen in CON due to better recognition, whereas the early deflection in IB and PL are due to not initially recognizing the target word, and so reobserving acoustic input to search for words with the best phonemic and semantic fit occurs.

As a summary conclusion of the analysis of our results, what was found is that context is relied upon more when more effort is needed for hearing. This has both positive and negative effects. It aids in improving accurate perception and retention of spoken language when predictions are accurate but can lead to false hearing when the information is misleading. The increased susceptibility to false hearing emphasizes the importance of the quality of auditory signals in choosing or designing areas where speech perception is fundamental to their success. Even in young adults with normal hearing there is still an increased risk of false hearing that has real world application in supporting designs to improve ease of hearing.

It is important to note the sample we have collected our data from only consisted of young normal hearing adults (ages 18 – 25), as there is a gap in literature which observes the neurophysiological properties of false hearing through EEG and the N400. Continuing research on older adults would be the logical next step in observing effects of aging on false hearing. Another reason for including older adults is, in theory, is to assure greater occurrence of FH+ trials (Failes et al., 2020; Rogers et al., 2012; Rogers & Wingfield, 2015), in turn providing more conclusive results on the differences in perception between FH+ and FH- trials. A larger sample size is also needed to make any conclusions on if false hearing occurs perceptually or post-perceptually overriding in favor of semantic fit. Determining which mechanism(s) drive the false hearing response would provide insight on cognitive systems and language perception more broadly and prove valuable in our understanding of the cognitive consequences of hearing impairment during speech perception.

REFERENCES

- Balota, D. A., & Spieler, D. H. (1999). Word Frequency, Repetition, and Lexicality Effects in Word Recognition Tasks: Beyond Measures of Central Tendency. *J Exp Psychol Gen*, *128*(1), 32-55. <https://doi.org/10.1037/0096-3445.128.1.32>
- Batterink, L., & Neville, H. J. (2013). The human brain processes syntax in the absence of conscious awareness. *J Neurosci*, *33*(19), 8528-8533. <https://doi.org/10.1523/JNEUROSCI.0618-13.2013>
- Benton, A. L. (1968). Differential behavioral effects in frontal lobe disease. In (Vol. 6, pp. 53-60). [https://doi.org/10.1016/0028-3932\(68\)90038-9](https://doi.org/10.1016/0028-3932(68)90038-9)
- Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *J Speech Hear Res*, *27*(1), 32-48. <https://doi.org/10.1044/jshr.2701.32>
- Blackford, T., Holcomb, P. J., Grainger, J., & Kuperberg, G. R. (2012). A funny thing happened on the way to articulation: N400 attenuation despite behavioral interference in picture naming. *Cognition*, *123*(1), 84-99. <https://doi.org/10.1016/j.cognition.2011.12.007>
- DEESE, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *J Exp Psychol*, *58*(1), 17-22. <https://doi.org/10.1037/h0046671>
- Elman, J. L., & McClelland, J. L. (1983). Exploiting lawful variability in the signal: The TRACE model of speech perception. *The Journal of the Acoustical Society of America*, *74*(S1), S68-S68. <https://doi.org/10.1121/1.2021096>
- Failes, E., Sommers, M. S., & Jacoby, L. L. (2020). Blurring past and present: Using false memory to better understand false hearing in young and older adults. *Mem Cognit*, *48*(8), 1403-1416. <https://doi.org/10.3758/s13421-020-01068-8>
- Gordon-Salant, S., & Fitzgibbons, P. J. (1997). Selected cognitive factors and speech recognition performance among young and elderly listeners. *J Speech Lang Hear Res*, *40*(2), 423-431. <https://doi.org/10.1044/jslhr.4002.423>
- Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: depth of retrieval. *Psychon Bull Rev*, *12*(5), 852-857. <https://doi.org/10.3758/bf03196776>
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, N.J. : Prentice-Hall.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol*, *62*, 621-647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203-205. <https://doi.org/10.1126/science.7350657>
- Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalogr Clin Neurophysiol*, *69*(3), 218-233. [https://doi.org/10.1016/0013-4694\(88\)90131-9](https://doi.org/10.1016/0013-4694(88)90131-9)
- MathWorks. (1990). Matlab. In.
- McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: downstream effects on older adults' memory for speech. *Q J Exp Psychol A*, *58*(1), 22-33. <https://doi.org/10.1080/02724980443000151>
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behav Res Methods Instrum Comput*, *36*(3), 402-407. <https://doi.org/10.3758/bf03195588>
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)

- Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behav Res Methods*, *47*(4), 1343-1355. <https://doi.org/10.3758/s13428-014-0543-2>
- Payne, B. R., Lee, C. L., & Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials. *Psychophysiology*, *52*(11), 1456-1469. <https://doi.org/10.1111/psyp.12515>
- Payne, B. R., Silcox, J. W., Crandell, H. A., Lash, A., Ferguson, S. H., & Lohani, M. (2021). Text Captioning Buffers Against the Effects of Background Noise and Hearing Loss on Memory for Speech. *Ear Hear*. <https://doi.org/10.1097/AUD.0000000000001079>
- Payne, B. R., Stites, M. C., & Federmeier, K. D. (2019). Event-related brain potentials reveal how multiple aspects of semantic processing unfold across parafoveal and foveal vision during sentence reading. *Psychophysiology*, *56*(10), e13432. <https://doi.org/10.1111/psyp.13432>
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., . . . Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear Hear*, *37 Suppl 1*, 5S-27S. <https://doi.org/10.1097/AUD.0000000000000312>
- Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *Q J Exp Psychol*, *20*(3), 241-248. <https://doi.org/10.1080/14640746808400158>
- Rogers, C. S. (2017). Semantic priming, not repetition priming, is to blame for false hearing. *Psychon Bull Rev*, *24*(4), 1194-1204. <https://doi.org/10.3758/s13423-016-1185-4>
- Rogers, C. S., Jacoby, L. L., & Sommers, M. S. (2012). Frequent false hearing by older adults: the role of age differences in metacognition. *Psychol Aging*, *27*(1), 33-45. <https://doi.org/10.1037/a0026231>
- Rogers, C. S., & Wingfield, A. (2015). Stimulus-independent semantic bias misdirects word recognition in older adults. *J Acoust Soc Am*, *138*(1), EL26-30. <https://doi.org/10.1121/1.4922363>
- Silcox, J. W., & Payne, B. R. (2021). The costs (and benefits) of effortful listening on context processing: A simultaneous electrophysiology, pupillometry, and behavioral study. *Cortex*, *142*, 296-316. <https://doi.org/10.1016/j.cortex.2021.06.007>
- Tun, P. A., & Wingfield, A. (1999). One Voice Too Many: Adult Age Differences in Language Processing With Different Types of Distracting Sounds. *J Gerontol B Psychol Sci Soc Sci*, *54B*(5), P317-P327. <https://doi.org/10.1093/geronb/54B.5.P317>