EVALUATING LISTENING EFFORT USING EVENT-RELATED BRAIN POTENTIALS

By

Amy Kemp

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Communicative Sciences and Disorders-Master of Arts

2016

ABSTRACT

EVALUATING LISTENING EFFORT USING EVENT-RELATED BRAIN POTENTIALS

By

Amy Kemp

Maximizing the efficiency of listening effort is a critical goal for hearing rehabilitation. The current study evaluates modulations in brain functions underlying semantic processing in easier versus more challenging listening conditions. Event-related brain potentials (ERPs) present an ideal methodology for evaluating changes in neural processes as a function of listening effort. Neural processes underlying semantics, indexed by the N400, were elicited by final words of sentences with higher compared to lower likelihood of completing a sentence. Participants were required to respond with longer (1000ms), middle (700ms), or shorter (400ms) response time deadlines (RTDs). Thirty-six typical adults completed the paradigm. Response accuracy was greater for longer RTDs compared to middle and shorter RTDs. Additionally, accuracy was higher for final words that were more predictable compared to words that were less predictable. Additionally, N400 peak latencies were longer for the middle and shorter RTDs compared to the longer RTD, specifically for the words that were less likely to complete the sentence. This approach utilizes an index of semantics, the N400, as the marker of LE. The delayed N400 response is thought to reflect the increased listening effort for more challenging conditions and may result from the increased cognitive demands associated with greater listening effort.

To my brilliant advisor, Amanda Hampton Wray.
Without whom this thesis would have never been completed,
And I would have never developed an addiction to coffee.

ACKNOWLEDGEMENTS

This work was supported in part by the Richard and Nancy Heiss Endowment and through the Michigan State University Department of Communication Arts and Sciences Research Fellowship. We would like to thank the members of the Brain Systems for Language Lab for their support in stimuli development and data analysis; Kraig Wray, Valerie Rose, and Laney Roehl. Additionally, we would like to thank David Eddins, Rahul Shrivastav, and Eric Hunter for their support and assistance in the development and interpretation of this research project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Listening Effort

Listening effort (LE) can be defined by the cognitive effort put forth towards hearing in meaningful environments (Pichora-Fuller & Singh, 2006; Gosseline & Gagne, 2010; Kiessling et al., 2003). LE can be exemplified broadly as two individuals with the same level of hearing sensitivity (including hearing impairment) whose performance is at the same level (e.g., percent correct) on a given listening task while one individual exerts much effort to achieve that performance level than the other who experience differing levels of ease of listening, and thereby different supra-threshold hearing abilities. LE can also be exemplified by the same individual who experiences different levels of listening ease in different scenarios. The reasons underlying these differences in hearing ability, or listening effort, are currently unclear.

Many prior measurements of LE have used questionnaires, rating scales, and self-reports, which rely on subjective reporting by the listener (Picou et al., 2013; Mackersie & Cones, 2011; Zekveld et al., 2011; Bernarding et al., 2013; Dawes et al., 2014). Quality of life performance measures have been found to reflect a broad spectrum of satisfaction levels with listening using hearing aids. The relationships between mental load and ease of listening have been measured using the Speech, Spatial, and Qualities of Hearing Scale (SSQ; Gatehouse & Noble, 2004). The SSQ measures listening difficulties across numerous domains, such as satisfaction, ease of listening, and difficulty with or without visual cues. Questions such as "Do you have to concentrate very much when listening to someone or something?" (Gatehouse & Noble, 2004, 98) were asked to quantify LE. Participants responded to questions using a 10-point scale, with the lower numbers indicating greater difficulty. The SSQ has been adapted for use as a "listening effort scale" (Dawes et al., 2014).

However, the most frequent measure of listening effort involves using a Likert scale that quickly asks a participant to rate their level of fatigue or effort put into listening, typically using a 5- or 7-point number scale. These subjective methods have been used with a variety of populations, including individuals with normal hearing and individuals with hearing impairment, from young to elderly. Across individuals who complete these questionnaires, whether with hearing impairment or normal hearing abilities, differences in the effort and quality of their listening skills are highly variable. This may reflect the differences in listening effort expended by these listeners, measurement variability, or both.

Subjective scales have increased our knowledge of the differing conditions in which variations in hearing abilities can occur and LE can be evaluated using these scales, which can be modulated and manipulated by condition (i.e. directionality, in noise, audiovisual, etc.). Participants with normal hearing demonstrate changes in LE as conditions increase or decrease in difficulty of hearing environments. However, individuals with hearing impairment have decreased performance levels along with a higher level of LE in more challenging listening conditions (Dawes et. al., 2014). This indicates that individuals with hearing impairment have significantly different measures of LE than individuals with normal hearing. Differences between populations with normal hearing and with hearing impairment reveal that the manipulation of listening conditions impacts performance and listening ability to a greater extent in persons with hearing impairment.

Other behavioral measures have also been used in LE research. These methods include single-task and dual-task paradigms. Dual-task paradigms that involve listening tasks in conjunction with either visual processing or memory task have been implemented in several studies that have attempted to identify an objective measure of LE. Dual-task paradigms involve having a subject complete two competing tasks, operating under the theory that the reduction in performance on the second task could be used to evaluate cognitive processes engaged in the first task (Desjardins & Doherty, 2013; Feuerstein, 1992). In these studies, the primary task is typically a listening task while the secondary task is a performance task. Reduced performance on the secondary task is then thought to reflect to LE. Downs (1982) conducted a study that implemented a dual-task paradigm used a listening task along with a visual/memory task, e.g. attending to a flashing light probe prior to giving answers. This study revealed delayed response times in the condition in which the participants were attending to the flashing probe. This delay was even longer in listening environments that were more demanding for the participant (Downs, 1982; Downs & Crum, 1987).

Dual-task paradigms have been used with individuals with hearing loss and have evaluated the person's ability to attend to auditory stimuli while performing visual/memory tasks. Hornsby (2013) assessed word recognition, word recall, and visual reaction times in an attempt to quantify LE. Hornsby hypothesized that fatigue was correlated to LE, such that as performance on the response time task decreased and listening conditions got harder, the level of required effort would increase. However, these results did not reveal significant differences between conditions; therefore, these results suggest that fatigue may be related to LE in less direct ways or ways that are not currently understood (Hornsby, 2013). Fraser and colleagues (2010) used a dual-task paradigm consisting of speech and tactile recognition tasks, and used accuracy and reaction time to measure LE. The authors found that speech recognition in noisy conditions was more effortful in the audiovisual modality than in the auditory modality alone.

This suggests that performing two tasks simultaneously increased cognitive load, and thereby increased LE (Fraser et al., 2010). Visual reaction task and word recognition paradigms were also used together to measure LE in individuals with HI, with conditions varying by visual cues and background noise (Picou, Ricketts, & Hornsby, 2013). Picou, Ricketts, & Hornsby (2013) measured effects of working memory capacity and verbal processing speed. Participants with hearing aids benefited in the aid of complex listening situations; however, individual processing speed was the factor that correlated with LE. These results demonstrate the need for additional assessment of individual processing speed and working memory abilities in order to better understand LE within the hearing impaired population (Picou et al. 2013).

Dual-task paradigms evaluating listening effort while completing a cognitive task (repetition or recall) simultaneously with an auditory task are prevalent in the literature currently. Specifically, target acoustic stimuli were presented in noise or in the absence of noise while competing stimuli were presented either as text, acoustically or audio visually (Hallgren, Larsby, Lyxell, & Arlinger, 2001; Larsby, Hallgren, Lyxell, & Arlinger, 2005). Dual-task paradigms between populations have provided a comparison of speed and accuracy between young/elderly individuals with normal hearing and young/elderly individuals with hearing impairment. An assessment between individuals with hearing impairment and absence of visual cues revealed a reduced performance. The addition of noise in the task proved more difficult, resulting in both reduced speed and accuracy for elderly participants and individuals with hearing impairment (Hallgren et al., 2001). The group with hearing impairment reported greater effort on both the cognitive and physical tasks. Larsby and colleagues (2005) further investigated the same task but adding differing noise conditions, finding the same results; the greatest difficulty, both in measured accuracy and in perceived effort, for individuals with hearing impairment was in noisy conditions.

The results from dual-task paradigms have suggested that LE increases when other factors come into play, such as noisy environments, visual distractors, and complex memorization tasks. However, a direct measure of LE is difficult to obtain from these tasks because of the requirement of performing two simultaneous tasks. As performance on both tasks decreases, the conclusions are confounded. Specifically, it is difficult to differentiate the ways in which the requirements of each dimension, or task, such as changes in listening requirements or demands on visual processing or memory capacity, are affecting performance. The secondary cognitive tasks, such as recognizing, discriminating, and disregarding irrelevant auditory probes, may require additional neural activity, which can limit one's ability to directly measure the primary task. One potential confound is that the secondary task may reduce performance effort available to engage in the primary task. Another potential confound is that, with additional tasks, the cognitive load increases, which could not only result in fatigue, but also alter cognitive processing in an unexpected, and potentially immeasurable, ways (Kramer, Trejo, & Humphrey, 1995; Ullsperger et al., 2001). While it is possible to infer information about LE from the performance on the secondary task (Sarampalis et al., 2009; Downs, 1982; Gosselin and Gagne, 2010; Desjardins & Doherty, 2013; Fraser et al., 2010; Hornsby, 2013; Picou et al., 2013), dual-task paradigms may not provide a direct evaluation of LE. Thus, given the potential confounds of dual-task paradigms, a single task paradigm, such as sentence completion task, would allow for more direct identification and measurement of LE (Allison & Polich, 2008; Harmony et al., 2000).

While behavioral studies have provided insights into LE, these types of studies also have limitations. Since listening is thought to be a skill, or cognitive effort, subjective measures of effort may not provide detailed or sensitive enough information to understand how and why listening effort varies so greatly across individuals. Subjective measures can have several confounds, e.g. subject bias, lack of introspection, and measures based on perceived performance or created excuses for lack of effort (Hicks & Tharpe, 2002; Sarampalis, Kulluri, Edwards, & Hafter, 2009; Fraser, Gagne, Alepins & Dubois, 2010; Mackersie & Cones, 2011). Limitations of behavioral measures are the correlation strength and variability between subjective and objective measures. In dual-task paradigms, the separation of efforts engaged in each task can be variable when participants subjectively identify their effort levels (Gosselin and Gagne, 2010; Desjardins & Doherty, 2013; Fraser et al., 2010; Hornsby, 2013). It is possible that LE is a multifaceted process and subjective and behavioral measures of LE provide insight into one aspect of LE. However, additional insight into LE can be gained from measures of physiology underlying LE.

Physiological Responses Underlying Listening Effort

More recent studies have applied physiological techniques to the measurement LE, including physiological measures such as heart rate, skin conductance, skin temperature, electromyographic activity (EMG), and pupil response (Zekveld, Kramer, & Festen, 2010; Mackersie & Cones 2011; Koelewijn, de Kluiver, Shinn-Cunningham, Zekveld, & Kramer, 2015; Fournier, Wilson, Swain, 1999; Kramer et al., 1997). Studies of heart rate have yielded inconsistent results, with some studies revealing correlations between increasing stress levels and increases in the difficulty of listening conditions (Fournier, Wilson, Swain, 1999; Kramer et al., 1997) while other studies reported no trends. However, skin conductance, skin temperature, and EMG studies have exhibited more consistent relationships with LE. Skin conductance, skin temperature, and EMG were all measured during an increasing letter recall task that required the listener to repeat the stimuli correctly. As the noise increased, increases in skin conduction, skin temperature, and EMG were observed (Mackerize & Cones, 2011; Fournier et al., 1999). The authors proposed that more difficult conditions required greater attentional resources, resulting in increases in stress response, which in turn resulted in the increases observed in physiological measures (Mackersie & Cones, 2011).

Pupil response has consistently been associated with increased LE (Koelewijn, de Kluiver, Shinn-Cunningham, Zekveld, & Kramer, 2015; Zekveld et al., 2010; Kramer et al., 1997). A study by Kramer (1997) found correlations between pupil dilation and LE performance when evaluating pupil dilation in conditions with 50% intelligibility either with or without noise. Increased task difficulty resulted in increased pupil dilation. Additionally, the participants with HI revealed less pupil reduction with increasing listening difficulty, along with more reported LE. Zekveld and colleagues (2010) asked participants to listen to sentences in conditions of 50%, 71%, and 84% intelligibility. Pupil dilation amplitude, peak latency, and mean pupil dilation were measured in each condition, along with subjective questions about LE. These comparisons revealed that pupil dilation increased with decreasing speech intelligibility (e.g., increased listening difficulty). Subjective measures of LE also increased, supporting that LE, as indicated by both questionnaires and pupil responses, increases with increasing task difficulty. Pupil response in both studies was hypothesized to change as a function of task difficulty, creating an objective measure of LE. A recent study by Winn and colleagues (2015) also investigated LE with regards to intelligibility and pupil dilation. Subjects were asked to listen to sentences and repeat back what they heard in order to measure intelligibility. Pupil responses were found to be sensitive to changes in listening performance. Differing amounts of pupil dilation between conditions in which the participant scored 100% revealed differing amounts of effort as the cognitive load required for comprehension increases.

Kuchinsky and colleagues (2013) investigated LE and pupil dilation in a relation to a word identification and response time measures in a population highly impacted by difficulties with LE, older adult participants who had hearing loss. The pupil response was found to be sensitive in all listening conditions, with a greater pupil size and sustained pupil size post stimuli. This finding supports the link between cognitive load and pupil response but indicates that further investigations are needed. Zekveld, Kramer, & Festen (2011) also studied pupil response in both HI and in middle-aged populations. Again, the participants listened to sentences in conditions of varying intelligibility and rated their level of LE. Participants also took part in a speech reception threshold task, a processing speed task, and a vocabulary test. The results from this study indicated that pupil response increased as the task became more difficult for all populations. However, for the middle-aged and hearing impaired populations, the tasks assessing cognitive load, processing speed tasks, and vocabulary tests indicated lower performance than for the normal hearing group. These findings suggest that the usage of pupil measurement may be more sensitive to differences between groups than other cognitive processing tasks, which may have other confounding factors.

However, these objective measures of a sensory response only provide indirect measures of LE and do not directly explain differences in ease of listening. Listening is a cognitive task that can be manipulated by an individual, in contrast to hearing, which is a sensory response that occurs with minimal cognitive control (Hicks & Tharpe, 2002; Mackersie & Cones, 2011). For example, even if the amplification of a sound increases, it does not change someone's ability to attend to and process the sound. Thus, LE is a cognitive function (Beck, 2011; Gosseline & Gagne, 2010; Pichora-Fuller & Singh, 2006). A more direct measure of cognition and cognitive control may be necessary to directly evaluate LE. Newer methodologies for measuring LE are now focusing on utilizing neuroscience techniques that more directly evaluate brain functions and the cognitive processes involved in understanding and processing speech.

Neural Processes Underlying Listening Effort

Only a handful studies to date have looked at the neural functions underlying LE using electroencephalography (EEG) (Bernading et al., 2013; Bertoli & Bodmer, 2014; Erlberk et al., 2014). EEG is a noninvasive neuroimaging technique that measures the changes in electrical responses from populations of neurons firing in synchrony (Nunez, 1995). EEG allows for real-time measurements of neural activity, enabling the evaluation of modulations of earlier, more automatic, as well as later, more cognitively controlled functions (Weisz & Obleser, 2014). Event-related brain potentials (ERPs) are EEG that is time-locked to a specific stimulus, providing information about neural activity related to processing the stimulus with exquisite temporal resolution (Nunez, 1995). ERPs can provide information about the neural processes underlying a myriad of cognitive functions, including potentially LE. ERP components such as the N100, N200, P300, and N400 have been used in LE research. The N100 and N200 are exogenous ERP components, which are modulated by the physical stimulus and can be observed whether the listener is attending or not attending to the stimulus. However, attention, or effort, can modulate the amplitude and/or timing of the response (Donchin et al., 1978). In contrast, endogenous components, such as the P300 and N400, are internally controlled responses, largely dependent on a person's intentions, actions, and response to the stimulus, representing a degree of cognitive control (Donchin et al., 1978).

N100 is an early automatic neural response elicited during both active and passive tasks. During active tasks, the N100 can be modulated by conscious effort and reaction to a task (Mulert, Menzinger, Leicht, Pogarell, & Hegerl, 2005). Studies of the N100 and LE have revealed an increase in activation with increasing task difficulty as a result of changes in task instructions, e.g., a relaxed condition and an effortful condition. Increased N100 amplitudes have been observed in tasks that require increased effort (Mulert et al., 2005; Matsuda & Nittono, 2014). Bernading and colleagues (2013) investigated the correlation between the N100, age groups, and degrees of hearing impairment. Increased N100 amplitudes with middle/old age suggested that there was increased attention during difficult listening tasks, but these changes did not meet significance. These findings, along with other studies (e.g. Grouppe et al., 2010; Straus, 2010; Pekkonen, Rinne, & Näätänen, 1995), suggest that in active paradigms, the N100 can be modulated by attention, which may play a role LE. The modulation of the N100 by tasks with increased LE supports the potential for using ERPs as a measure of LE in other aspects of cognitive processing, such as language (Shahin, 2012; Straus, 2010).

N200 is elicited by a stimulus that is unpredicted in a sequence of predictable events, such as a deviant stimulus in an oddball paradigm (Pritchard et al., 1991, Patel & Azzam, 2005). The N200 is thought to be related to inhibitory control, which is the ability to attend to one stimulus while another stimulus is occurring simultaneously. This suggests that the N200 is part of a larger cognitive processing system controlled by attention, specifically auditory sensory memory (Patel & Azzam, 2005; Erlbeck et al., 2014). The amplitude of the N200 has been found to increase as a function of instructional condition, e.g. ignoring or attending to the stimuli, which affects the quantity of task specific errors and perhaps is an indication of perceived effort (Benikos et al., 2013; Erlbeck et al., 2014). Other studies have found that, in both young children populations and aging populations, N200 amplitudes decrease with changes in task instruction. This is posited to result from an immature or decreased cognitive processing system that restricts the attention given to the auditory sensory memory (Patel & Azzam, 2005; Pekkonen, Rinne, & Näätänen, 1995; Amenedo, Diaz, 1998; Humes & Wilson, 2003; Humes & Christopherson, 1991).

P300 is an endogenous ERP response has been posited to serve as a measure of LE, as P300 is thought to be involved in working memory and memory updating (Donchin & Coles, 1993; Patel & Azzam, 2005; Polich & Herbst, 2000). As demonstrated in the aging population, decreases in P300 amplitudes correlate with reduced speech recognition task accuracy for different conditions of listening difficulty (Polich & Herbst, 2000). While listening to a story in noise, (Kramer, et. al, 2005) participants were asked to respond to an oddball tone during a listening program. Questions were asked post session to determine understanding of story spoken in paradigm. P300 amplitudes decreased with increasing task difficulty, and performance on the comprehension task also decreased. However, the ability to isolate LE in the specific task used in this study was not clear (Kramer et al., 1995). Another study modulated a listening task difficulty by requiring a faster response time. The more difficult condition resulted a later onset of the P300. This later onset has also been tied to the predicted LE measure, but again, it was not clear whether LE alone was being measured (Strauss, 2010; Polich & Herbst, 2000). Bertoli & Bodmer (2014) aimed to evaluate changes in LE in older adults using a listening task with increasing difficulty by measuring changes in P300 amplitudes. This measure captured the increased effort and correlated self-reported stress of the participant, but did not reveal significant changes in P300 amplitudes as a function of listening difficulty. Together, these studies using ERPs to evaluate LE have demonstrated that neurophysiological measurements are a promising direction for evaluating LE.

The N400 is an endogenous ERP component, meaning that it functions independent of the stimulus parameters and reflects aspects of cognitive processing. The N400 is believed to reflect semantic processing and is thought to index ease of lexical access and integration, including retrieval from semantic memory and integrating semantic information within previously established contexts (Kutas & Federmeier, 2000; 2011; Lau, Phillips, & Poeppel, 2008). The amount of attention and automatic function given to processing a word may also be reflected by this ERP component (Kutas & Hillyard, 1980; Kutas & Federmeier, 2000). The difficulty of lexical integration of a stimulus modulates the amplitude, and potentially the timing, of the N400 in single word, sentence, and discourse contexts (Groppe et al., 2010; Kutas & Hillyard, 1984; Sivonen et al., 2006).

DeLong, Urbach, & Kutas (2005) investigated the activation process in incoming sentences. Manipulations of the interpretation of the word or sentence elicited a late N400. This indicated that the neural responses that are represented by the N400 are involved in decoding a sentence, playing a role in to comprehension. Sivonen and colleagues (2006) replaced initial phonemes with coughs in high and low probability stimuli. The participants responded verbally, producing the word that had been modified. The authors suggested that later N400 onsets in response to highly unexpected words resulted from a delay in word integration. Groppe et al. (2010) expanded upon the study by Sivonen and colleagues (2006) by alternating where a disrupting auditory stimulus (e.g., cough) was placed within a word. Their goal was to investigate the ways in which adults processed the remainder of the sentence after the distracting stimulus was presented. The results revealed a delayed N400 response elicited by a disruption in a sentence, which is believed to be associated with slower processing of the critical word (Groppe et al., 2010). A disruption in utilizing the sentence context to interpret the message creates a more challenging condition in which listeners can comprehend sentence-level information (Groppe et al., 2010; Sivonen et al., 2006; DeLong, Urbach, & Kutas, 2005). Van de Meerendonk, Kolk, Vissers, & Chwilla (2010) evaluated the ways in which perception changes when words are easier or harder to predict. The N400 was elicited when there was an unexpected event (i.e., the end of the word was hard to predict), but the word could still be integrated into the sentence. The N400 was delayed in onset if there was manipulated plausibility, meaning it was less likely that a given word was acceptable. Erlbeck and colleagues (2014) used a dual-task attention paradigm to manipulate effort, with participants either passively or actively attending to a stimulus. Participants completed a visual task, attending to a specific scene, or they passively listened to stimuli while focusing on a fixed target. Additionally, participants actively listened and responded to errors made in tones, words, or sentences. After each task, the participants subjectively rated their effort. In the active compared to the passive tasks, for both visual and auditory stimuli, decreased N400 amplitudes were observed.

Subjective measures reflected no change between the active and passive tasks. Despite changes in N400 amplitude as a function of engagement in the task, it difficult to know exactly which task manipulations resulted in changes in the ERP component.

The Current Project

This current project aims to evaluate LE using a single task approach to develop a reliable, quantitative measure of LE utilizing ERPs, focusing on neural processes underlying semantics to evaluate changes in LE. Developing a technique to quantify LE using a neurophysiological methodology, specifically ERPs, will enhance our understanding of LE as well as contribute to our ability to develop a reliable measurement of LE for use in clinical practice. The N400, an ERP component reflecting ease of semantic integration, can be modulated by word frequency and the difficulty level of integrating a word into a sentence. One long-term goal of this line of research is to identify neural components and paradigms that can be used to index changes in LE (Groppe et al., 2010; Kutas & Federmeier, 2000). In the current study, we will use the well-established ERP component, the N400, to provide an index of LE (Figure 1).

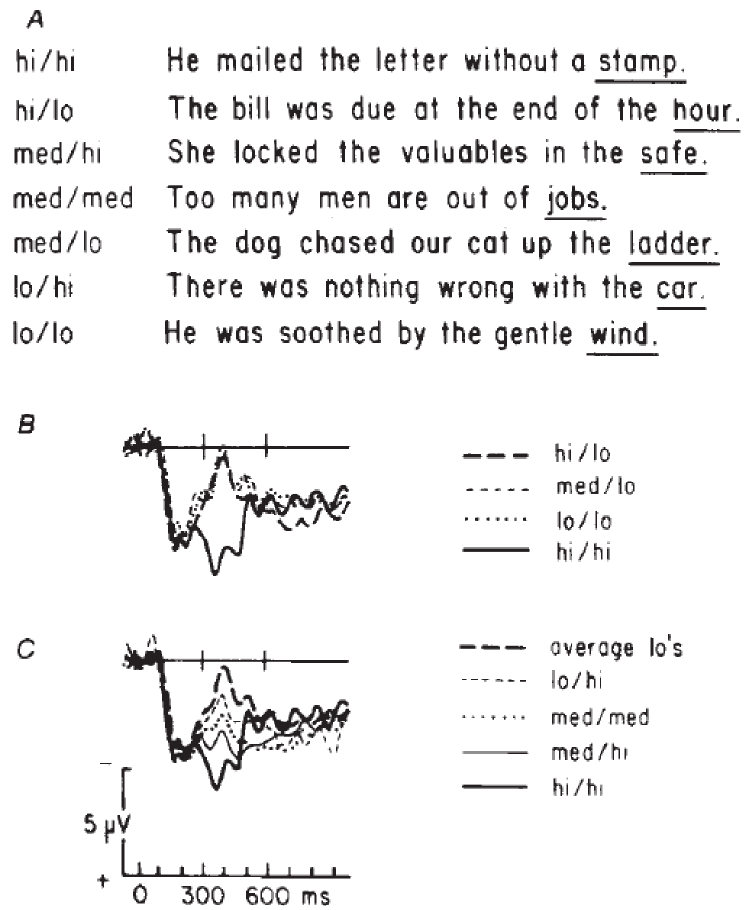


Figure 1. N400 amplitudes vary as a function of likelihood of a target word completing a sentence (Kutas & Hillyard, 1984)

In the current study, participants listened to sentences presented in the presence of four-talker babble noise with an individualized 70% signal-to-noise ratio (SNR). Sentences presented had either a high cloze probability final word, words with a high probability of completing the sentence, or the final word had a low cloze probability, words with a low probability of completing the sentence. Participants identified the final word in a sentence in a four-alternative forced choice task. Participants had three reaction time deadlines, a Long (1000ms), Middle (700ms) and Short (400ms) period of time in which to respond. Varied reaction time deadlines have been used in previous studies using pure-tone paradigms to evaluate differences in LE by varying the response demands in different listening conditions.

Results of this previous study are illustrated in Figure 2 (Benikos et al., 2013). Benikos and colleagues (2013) revealed that conditions with more difficult, or shorter, response time deadlines (Figure 2 High) elicited ERP components, P300s, with larger amplitude compared to conditions with easier, or longer, response time deadlines (Figure 2 Low).

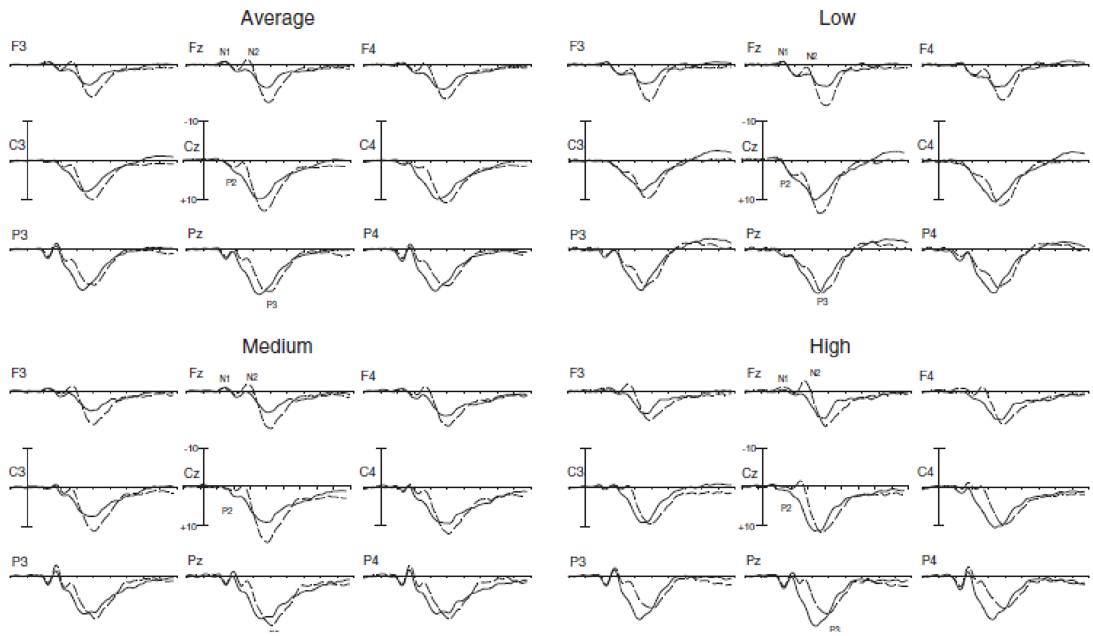


Figure 2. ERP response (P300) elicited by different response time deadline (RTD) conditions. Low task difficulty was a RTD of 1000ms, Middle difficulty was a RTD of 500ms, and high difficulty was a RTD of 300ms (Benikos et al., 2013)

The current study manipulates the difficulty of the listening task in two dimensions, probability of hearing a word, e.g., more (easier) or less (harder) likely, or reductions in response time deadlines, e.g., longer (easier), middle, or shorter (harder) time to read response options. Based on previous studies of LE, we hypothesize we will see decreased performance accuracy with increased cognitive demand. Specifically, conditions with Low close probability endings or shorter RTD will result in lower accuracy than High cloze or longer RTD conditions. Based on previous research on the N400 and in LE, we hypothesize that as the difficulty of the listening task increases, N400 amplitudes will also increase, both for high versus low cloze probability words and also for longer versus shorter reaction time deadlines. We predict that this change in N400 amplitude will reflect the increased effort used to integrate a word into a sentence context, either as a function of the likelihood of hearing a specific word, or the time pressure to respond. Thus, we posit that when an individual is dedicating greater amounts of LE to performing a task, N400 amplitudes be larger, reflecting increased engagement of neural processing.

METHOD

Participants

Sixty-three participants between the ages of 18 to 39 years old (mean= 22.9, SD= 3.86) were recruited. They were recruited from MSU SONA Paid Research Site and MSU campus. Based on self-report, adults with history of neurological injury (ADHD, epilepsy, concussion, or traumatic brain injury) and adults taking psychoactive medications were excluded from the present study. All participants included in the ERP analyses were right handed, monolingual speakers of English with normal or corrected-to-normal vision. All participants had hearing thresholds 20 dB HL or below at octave frequencies between 500 and 8000 Hz (ANSI, 2010), and specific hearing thresholds for each participant were obtained. From a total of sixty-three participants, ten were participated as pilot subjects, fifteen were excluded due to excessive EEG artifact (determined by fewer than 10 trials per condition), and two were excluded due to technical difficulties. The final group included 36 participants (21 females; 15 males). Of those 28 participants (mean age=23.0, SD=4.52) reported to be white/Caucasian, four were African American, two were Asian, and two chose not to report race.

Participants completed a background questionnaire which gathered information on education level and profession of the participant and parents/caregivers. Of the final sample, the majority of participants was in their third or fourth year of an undergraduate degree (n=18) or had completed a bachelors degree (n=8). The education factor was calculated for all participants on a seven point scale (Hollingshead, 1974). The education factor was calculated by the following criteria: persons with an education level less than 7th grade received a 1, persons who completed through 9th grade received a 2, persons who completed 10th or 11th grade received a 3, persons who completed high school or equivalent received a 4, persons who completed partial college (at least one year) or received special training received a 5, persons who completed standard college received a 6, and persons who complete graduate professional training received a 7. Mean (*SE*) participant education was 5.22 (*0.11*), indicating that the average education level was partial college. The maternal education, education level of participants mother's, which is a proxy variable for socioeconomic status (SES) for all participants was calculated utilizing the same seven point scale (Hollingshead, 1974). Across participants, mean (*SE*) maternal education was 5.36 (*0.18*), revealing that the average maternal education level was also partial college.

Behavioral Assessments

All participants completed a series of behavioral tasks, including the Test of Nonverbal Intelligence (Brown, Sherbenou, & Dollar, 1982), the Test of Adolescent and Adult Language (Hammill, Brown, Larsen, & Wiederholt, 1994), and Stroop: Color and Word Test (Golden, Freshwater, & Golden, 2003), to evaluate executive function, specifically inhibition, skills. The Stroop task was designed to evaluate the ability to inhibit a prepotent response in order to produce a more appropriate response. The individual suppresses the reading response of word naming and instead produces the name of the color font. This score demonstrates the ability to selectively sort and react to information in one's environment (Stroop, 1960). All participants had normal nonverbal intelligence, language, and executive functioning. All participants also completed an ERP task, described in detail below. Behavioral measures and ERPs were collected in 2.5-3 hour one session. All participants were paid for participation. Informed consent was obtained from all participants and study procedures were approved by the Michigan State University Human Research Protection Program (HRPP).

ERP Stimuli

ERP stimuli consisted of 300 sentences that are syntactically and semantically correct with high cloze (easier) or low cloze (harder) probability final nouns (target words) (Krause, Kennedy, & Nelson, 2014; Kutas & Hillyard, 1980). The sentences were 6-11 words in length. The sentences were presented at an average of 69-72 dB. Each sentence base (N = 150) contained a high cloze probability and low cloze probability ending, i.e. “I only eat toast with lots of grape *jelly*” versus “I only eat toast with lots of grape *juice.”* Each word that served as a high cloze probability ending for one sentence also served as a low cloze probability ending for a different sentence, so that the high and low cloze probability word lists were completed counterbalanced. The complete list of sentences is presented in Appendix A. All sentences were developed using simple sentence structures. All target words are words commonly known by children aged five years and younger, and are words included on the MacArthur Communicative Development Inventory (CDI; Fenson et al., 1993) and/or Spoken Word Count frequent words used by five-year-olds (Wepman & Hass, 1969).

All sentences were tested on thirty-one young adults who did not participate in the ERP portion of the experiment in order to determine the cloze probability of each sentence-final target word. These subjects were given a survey with the last word omitted and asked to finish the sentences with the first word that came to mind (Hagoort & Brown, 2000). Based on the cloze probability, a set of 150 highly likely, or high cloze probability, sentences were selected with a mean cloze probability of 60% (range 30-100%). A matched set of 150 sentences (same base sentences) with unlikely, but semantically acceptable, or low cloze probability endings were also developed, with a mean cloze probability of 0.78% (range 0-30%; Hagoort & Brown, 2000).

All sentences were masked by four-talked babble. For each participant, SNR was calculated using a program that calculated an individual's 70% SNR. A two-down, one-up program was used to determine the SNR at which the individual could identify sentence-final words with 70% accuracy. Participants heard thirty low cloze probability sentences presented in noise. None of the SNR paradigm sentences were used in the main ERP portion. Participants identified the final word in a four-alternative forced choice task. The noise level was adjusted using the following criteria: If the participant answered two trials correctly in a row, the noise level increased by 1 dB. If the participant answered one trial incorrectly, the noise level was turned down by 0.5 dB. The noise level at which the participant could respond with 70% accuracy was the attenuation level at which the four-talker babble was presented throughout the ERP experiment. The average attenuation level was 0.045 dB (SD=0.0187).

Stimuli were recorded by a native English speaker with a neutral American accent using headset microphone and digital Linear PCM Recorder at 44100 Hz. The onset of each word in each sentence was determined by visual inspection of the sound waveform by two independent researchers using Praat. Any discrepancy in timing of the word onset was resolved by a third researcher. Sentences were presented in blocked order by reaction time deadline (see below). Sentence presentation order was pseudo randomized, such that no base sentence was repeated within one block and no more than three high or low cloze probability final words were presented consecutively.

At the end of each sentence, participants identified the final word of the sentence from a visual four-alternative forced choice task. Four phonetically similar words, all simple nouns, were presented, with the location of the correct response pseduorandomized between sentences and blocks. Task difficulty was modulated by manipulating an amount of time the response choices were available on the screen following to each sentence, known as a reaction time deadline (Benikos et al. 2013). Three reaction time deadlines (RTDs) were used; the Long (1000ms), Middle (700ms), and Short (400ms) RTDs. Decreasing the time available to read and select the correct response increased the response pressure, thus requiring the participant use more effort to complete the task. RTDs were blocked such that participants knew how much time they had to respond for each sentence. Each block consisted of 25 sentences. Three block presentation orders were created, with block order pseudorandomly assigned. Each block order began with a Long condition and no RTD repeated more than twice in a row. Participants were assigned a block presentation order prior to arrival.

Procedure

After providing written consent for participation and placement of the electrode cap, participants were seated in a sound-attenuating booth. The sentences were presented auditory through ER-1 insert earphones. A computer monitor was positioned approximately 145 cm in front of the participant. Before the data was recorded, participants were provided these instructions for the calibration portion, which also served as the training sequence:

"For this part of the study, your hearing level in noise will be found. You will hear 30 sentences in continual noise. The noise will sound like you are at a busy restaurant or party. For each sentence, a "Ready?" will appear on the screen. Press any button to advance when you are ready. Next, a crosshair will appear on the screen. While the crosshair is on the screen, stare at the crosshair and the sentence you are listening for will be played. Then, 4 words will appear on the screen. These words are arranged similarly to your answer pad with the top word being coordinated with the top button, left with left, and so on. Pick the word you heard as the last word in the sentence. The words may be real words or nonwords, they might fit well in the sentence, or they might not. During this part, the sound level of the noise will increase and decrease. If you cannot read the words fast enough, no other words, such as the next "Ready?" will appear until you select an answer, so go ahead and choose a response to the best of your knowledge/memory."

Once the participant was ready, the four-talker babble began. After about 10,000ms, a "Ready?" would appear on the screen, and he/she pressed a button to begin the sentence. Then a fixation crosshair appeared on the screen. At the completion of each sentence, participants selected the sentence-final word they heard from a four-alternative forced choice task. After responding, the word "Ready?" appeared on the screen and the participant pressed a button to begin the next trial/sentence. The four-talker babble was played throughout the task, increasing and decreasing based on the participants' response. After the calibration task was administered, an attenuation score was obtained. This was the SNR that was utilized throughout the main portion of the experiment, during which ERP data was acquired. For the main experiment, the participant was provided these instructions:

"For this part of the study, you will be asked to do a similar task as you did before. You will hear 12 different blocks with 25 different sentences; this will take about an hour to complete. There will still be the continual noise in the background, but this time it will stay the same level. Again, for each sentence, "Ready?" will appear on the screen. Click any button to advance when you are ready. Then a crosshair will appear on the screen. While the crosshair is on the screen, stare at the crosshair and the sentence you are listening for will be played. Four words will then appear on the screen. These words are still arranged just as your answer pad is. Again, choose which word was the last word you heard in the sentence. The words may be real words or nonwords; they might fit in the sentence, they might not. After you pick your word, you will be given feedback as to whether or not you chose the correct word. Try to get as many right as you can.

For each block there is a differing time response to choose the word you hear. Before each block, there is a time stamp telling you how long the words will be displayed on the screen. Timing varies from Short about 400ms, Middle, about 700ms, and Long, 1000ms. If you do not answer before the words disappear, the program will not continue until you choose an answer. Pick a response that you think fits best. Do you have any questions"

Once the participant was ready, the four-talker babble began. Similar to the calibration portion, a "Ready?" would appear on the screen to start each trial. The participant would press a button to advance the paradigm and the fixation crosshair appeared on the screen. A sentence would be played through the headphones. At the completion of each sentence, participants selected the sentence-final word they heard from a visual four-alternative forced choice task. Real-time accuracy feedback was provided to encourage participants to maintain high levels of effort throughout the paradigm. After responding, the word "Ready?" appeared on the screen and the participant pressed a button to begin the next trial/sentence. The four-talker babble was played throughout each block. This continued for the remainders of the blocks with a 5-10 minute break being given after block eight. Figure 3 demonstrates the layout of the paradigm from the perspective of the participant.

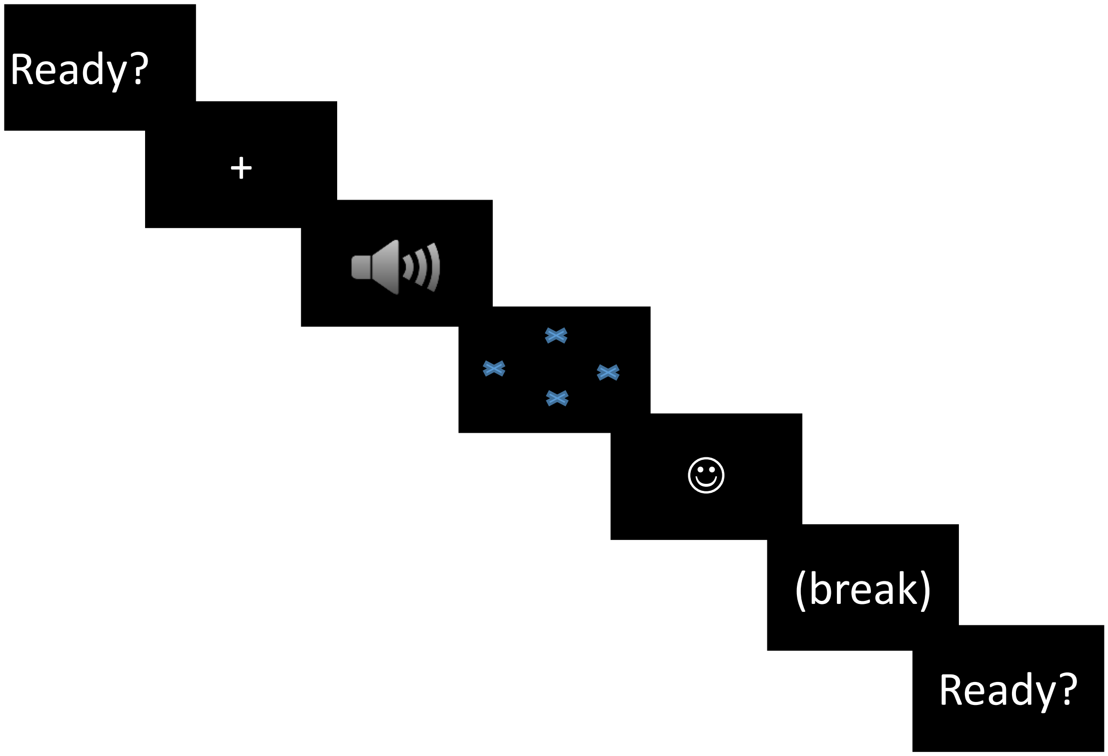


Figure 3. Paradigm procedure from participant perspective

Data Acquisition and Analysis

EEG data was collected using an elastic cap with 32 Ag-Ag/Cl electrodes embedded in it (Biosemi, Amsterdam, Netherlands). Electrode locations were consistent with the International 10-10 system (Figure 4). Additional electrodes were placed over the left and right mastoids, and to monitor eye movements, electrodes were also be placed over the left and right outer canthi, and below the left orbital ridge. EEG data was recorded at 512 Hz relative to a Common Mode electrode. Offline, EEG data was down sampled to 256 Hz and referenced to the average of electrodes placed over the left and right mastoids.

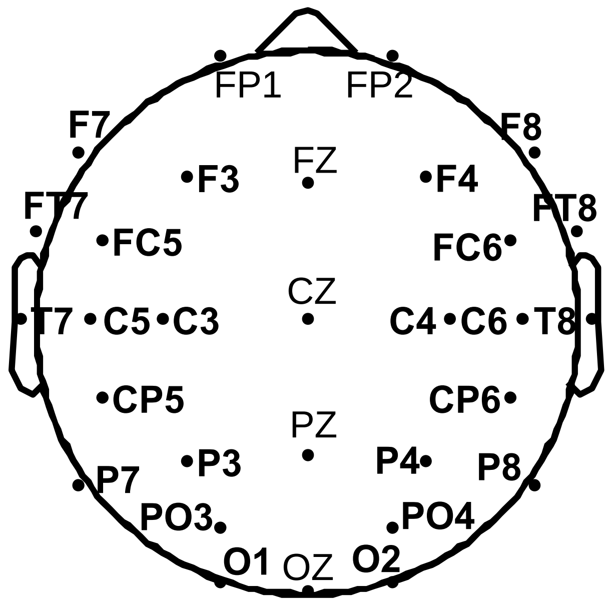


Figure 4. Electrode placement

ERP analyses were conducted using EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014). Data was band-pass filtered from 0.01 to 40 Hz and epoched between 100ms prior to and 1500ms after target word onset, using the first 100ms as the pre-stimulus baseline for correction. Only trials/epochs in which the participants selected the correct response were included in analyses. Automatic artifact rejection was initially executed using a 200ms window that moved at 50ms increments. Ocular data that exceeded a change of 100 �V within the time window, or other data that exceeded a change of 200 �V within the 200ms time window were automatically marked as artifact. Following automatic artifact rejection, a subsequent epoch-by-epoch manual artifact rejection step was performed by the author to exclude any remaining epochs containing eye movements and muscle artifacts. Additional epochs were rejected if the epoch contained eye movements and/or muscle artifact. Out of ~150 trials per condition, a mean (*SD*) of 62.08 (*26.61*) trials per participant were accepted for the high cloze condition, and 55.94 (*22.27*) trials per participant were accepted for the low cloze condition. Although this number of accepted trials is low, true ERP data was believed to be captured with the remaining trials.

Statistical Analysis

N400 mean amplitudes elicited by the final word in each sentence were measured within an earlier time window, between 300 to 600ms post-stimulus onset, and a later time window, 530 to 800ms post-stimulus onset. N400 local peak latencies elicited by the final word in each sentence were measured as the most negative point with at least 4 more positive points on each side between 100 to 800ms post-stimulus onsets. Twenty-four subjects had EEG recordings that included faulty electrodes. In this situation, the measurements for mean amplitude and peak latency for the faulty electrode were replaced with the value of the mean across all other participants at that electrode location.

Behavioral data of accuracy and reaction time were analyzed using repeated-measures ANOVAs with within-subject factors of reaction time deadline (Long, Middle, Short) and cloze probability (High, Low). ERP data was analyzed using repeated-measures ANOVAs with within-subject factors of reaction time deadline (Long, Middle, Short), cloze probability (High, Low), hemisphere (Left, Right), electrode location (Frontal, Frontocentral, Central, Centroparietal, Parietal, Occipital), and laterality (Lateral, Mid-lateral). Further analyses were performed on interactions for which $p < 0.05$ in order to isolate significant interactions. For the purpose of this project, the interactions between RTD and cloze probability were analyzed.

RESULTS

Behavior

Reaction Time

Mean (*SE*) reaction time across all conditions was 1180.66 (*33.69*)ms post-stimulus onset. Per condition, average reaction times (*SE*) were as follows: Long RTD – 1201.99ms (*32.71*), Middle RTD – 1162.65ms (*33.32*), Short RTD – 1171.50ms (*43.92*). There were no differences in reaction time between the high and low cloze probability sentences (Cloze: F (1, 35) = 1.32, p=2.58). There were also no significant effects of RTD or interactions between the three reaction time deadlines and cloze probability (RTD & RTD X Cloze: F (2,70) < 1.62, p > .208).

Performance Accuracy

Across all three conditions, performance accuracy decreased from the Long RTD to the Middle RTD to the Short RTD (F (1, 35) = 94.91, p < .001, n_p^2 = .731). Overall mean (*SE*) accuracy across all three RTD conditions was 82.1% (*0.015*). Mean (*SE*) accuracy for the Long RTD was 85.0% (*1.98*), decreasing to 79.1% (*1.39*) for the Middle RTD, and to 78.1% (*2.02*) for the Short RTD. Participants were more accurate detecting the high cloze compared to the low cloze probability words (Cloze: F (1, 35) = .001, p = .174). There was a significant interaction between RTD and cloze probability, F (1, 35) = 7.36, p = .008, n_p^2 = .174. Step-down ANOVAs were conducted to clarify the interactions between accuracy and RTD. The interaction between the Long RTD and Short RTD was significant, RTD X Cloze: F (1, 35) = 5.22, p = .029, n_p^2 =.130. Additionally, the interaction between the Middle RTD and Short RTD was significant, RTD X Cloze: F (1, 35) = 7.83, p = .008, n_p^2 = 184. The interaction between Long RTD and Middle RTD was not significant, RTD X Cloze: F (1, 35) = 1.35, p = .254, n_p^2 =.037. Figure 5 illustrates the interactions between performance, RTD, and Cloze condition.

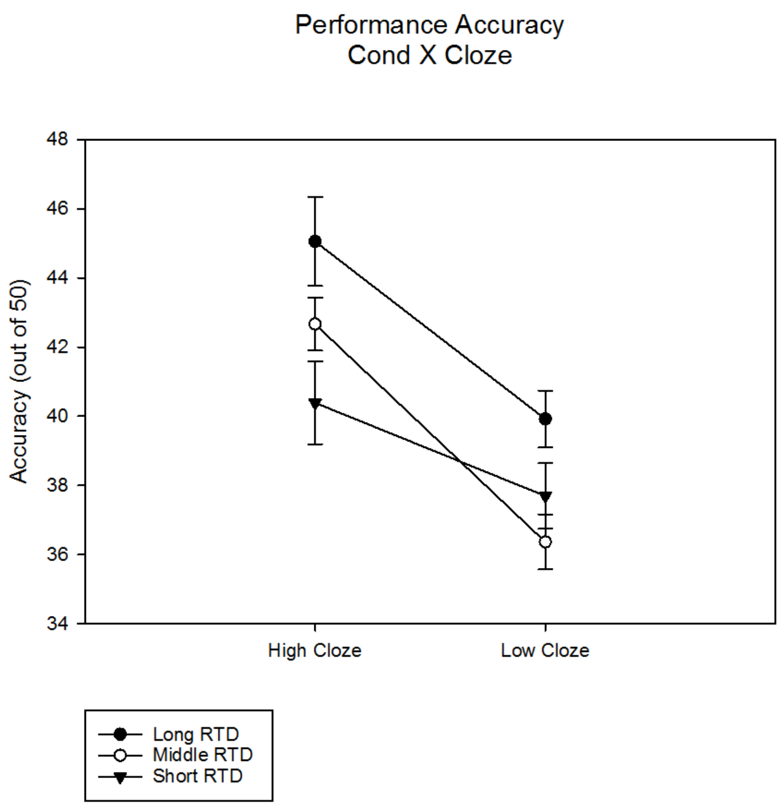


Figure 5. Peformance Accuracy between RTDs and Cloze probabilities

N400

Mean Amplitude

Grand average waveforms for the three RTD conditions are illustrated in Figures 5, 6, and 7 for the Long RTD, Middle RTD, and Short RTD, respectively. There was a significant condition effect of cloze probability in the 300-600ms time window. Low cloze sentences elicited larger N400 mean amplitudes than High cloze sentences (Cloze: F (1, 35) = 5.38, p = .026, n_p^2 = .133). Differences between High and Low cloze conditions are illustrated in the grand average ERP plots for each RTD condition (Figures 6, 7, and 8). In the same time window, there were no significant effects or interactions involving RTD or Cloze condition for the N400 mean amplitudes measured (RTD & RTD X Cloze: all F (2, 70) < 1.10, p > .304).

Within the 530-800ms post-stimulus onset time window, a trend toward a difference between cloze conditions was observed (Cloze: F (1, 35) = 3.55, p = .068, n_p^2 = .092), with the Low cloze condition eliciting larger N400 amplitudes than the High cloze words. As in the earlier time window, there were no significant interactions involving RTD or Cloze Probability (RTD & RTD X Cloze: all F (2, 70) < 3.56, p > .0.68). Although the grand average plots appear to reveal emerging N400 mean amplitude differences across cloze probability conditions, no significant effects were observed between mean amplitudes across conditions.

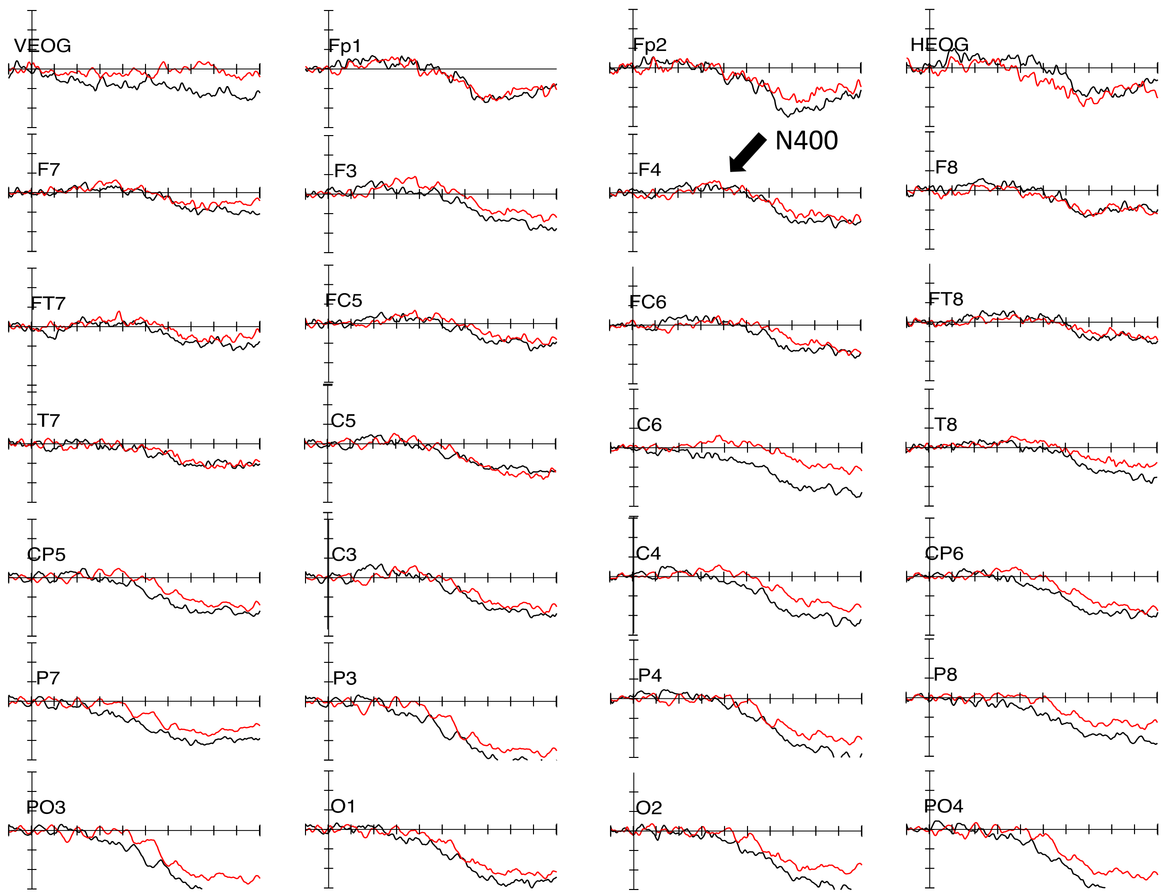


Figure 6. Long RTD ERP responses shown above (n=36). See below for key

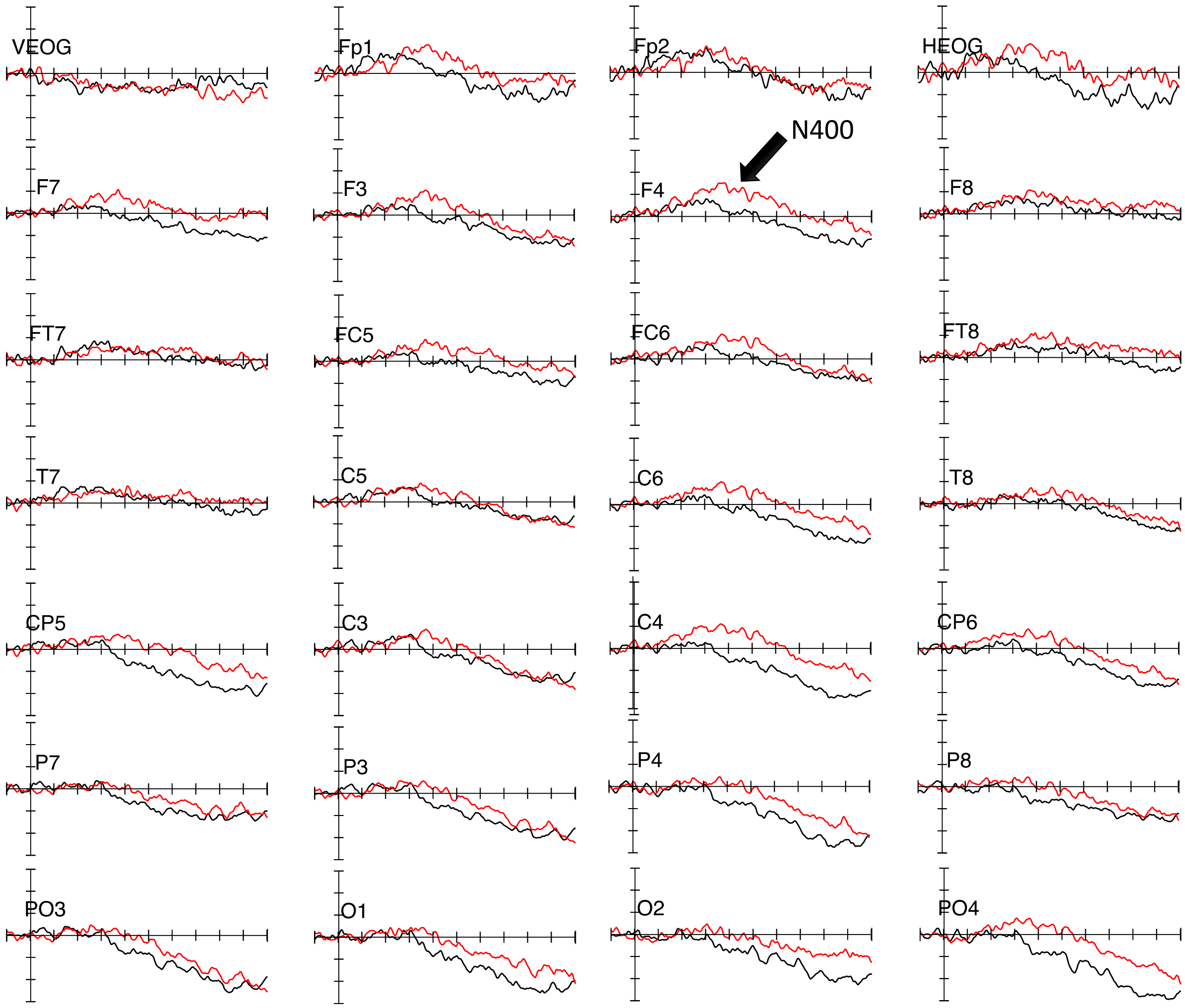
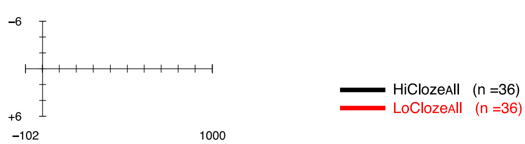


Figure 7. Middle RTD ERP responses shown above (n=36). See below for key



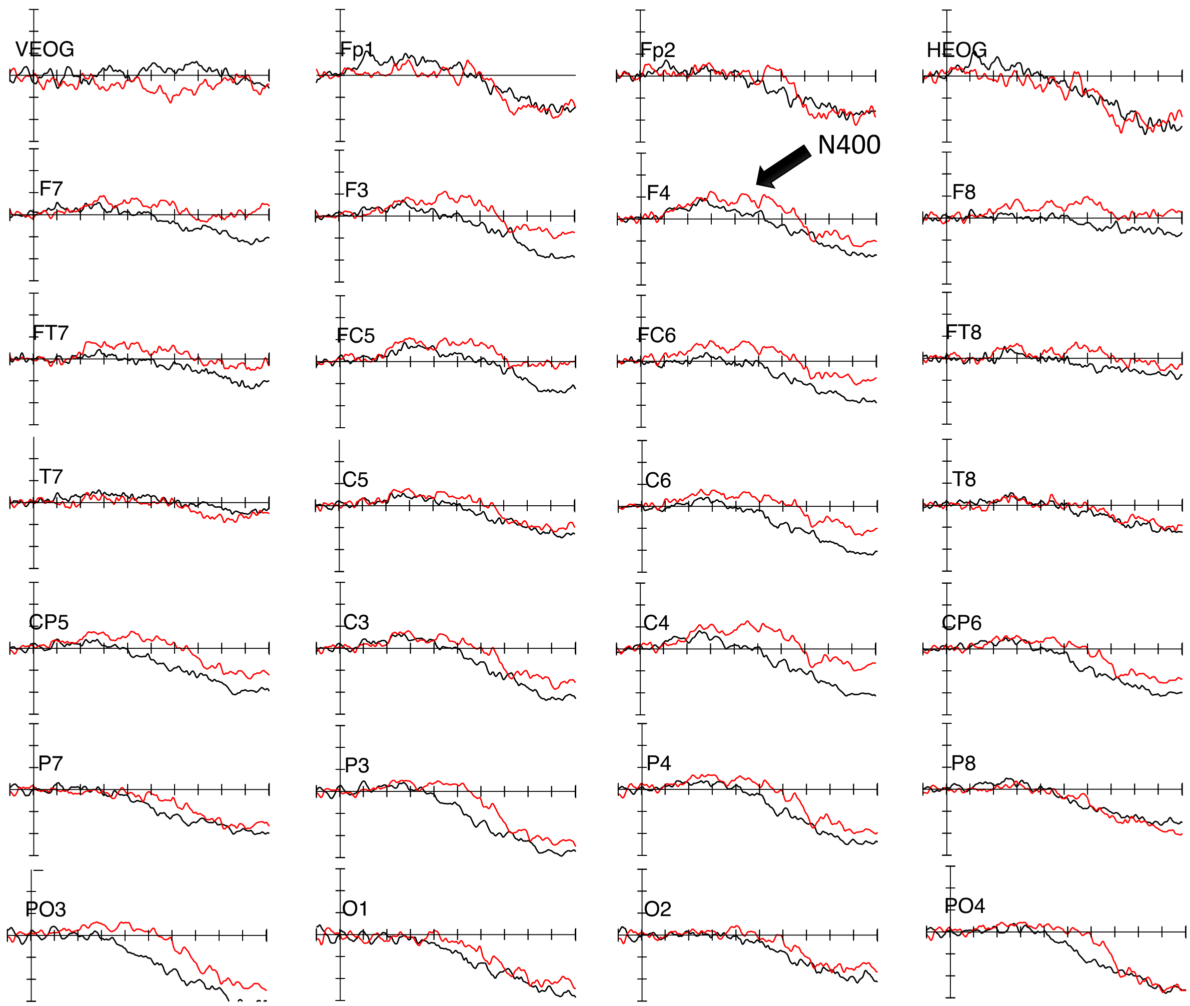
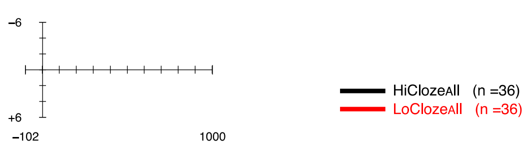


Figure 8. Short RTD ERP responses shown above (n=36). See below for key



Peak latency

N400 peak latencies were measured between 100 and 800ms post-stimulus onset. N400 peak latencies were earlier for the High compared to the Low cloze words (Cloze: F (1, 35) =17.39, $p < .001$, n_p^2 =.332). An interaction between RTD, cloze condition, and hemisphere revealed differences in N400 peak latencies between conditions, F (1, 35) =6.68, p = .002, n_p^2 =.160 (Figure 8). To further evaluate this interaction, step-down ANOVAs between two RTD conditions were performed. Comparisons between the Long and Short RTDs revealed an interaction between RTD, Cloze probability, and hemisphere (RTD X Cloze X Hemi: F (1, 35) =15.93, p < .001, n_p^2 =.313), with the Long RTD earlier than the Short RTD over the right hemisphere, which included all electrode locations. Comparisons between Long RTD and Middle RTD revealed an interaction between RTD, Cloze probability, and Hemisphere (RTD X Cloze X Hemi: F (1, 35) =7.36, p = .010, n_p^2 =.174). Comparisons between the Middle and Short RTDs revealed a no significant interaction of RTD, Cloze probability, and hemisphere, F (1, 35) =.213, p = .648, n_p^2 =.006, demonstrating the significant interaction between Low cloze and shorter RTDs (Middle and Short), which can been seen in Figure 9.

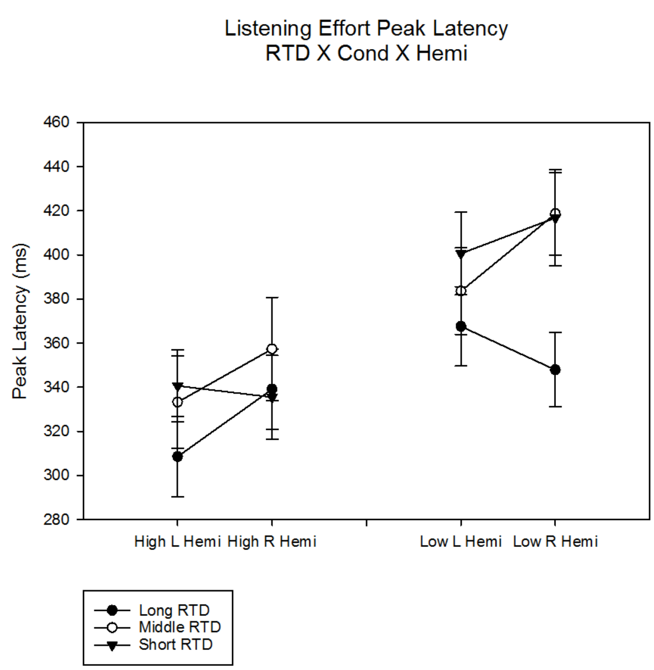


Figure 9. Depiction of interaction between RTD and Cloze probability Conditions

DISSCUSSION

The current study aimed to evaluate the effects of listening effort on neural processes underlying semantics by assessing changes in the N400 component as a function of cognitive demand, or effort, in challenging listening environments. N400s were elicited by either high cloze (highly likely) or low cloze (less likely) probability words at the end of sentences presented in noise (four-talker babble). Participants identified the final word of each sentence in a visual four forced-choice task. Cognitive demand for the task varied by reaction time deadline; participants saw the four reaction choices for a long period of time (Long, 1000ms), and shorter time period (Middle, 700ms), or a brief period (Short, 400ms). Behavioral and neurophysiological responses were evaluated as a function of the three reaction time deadline conditions.

Participants responded with greater accuracy for the high compared to the low cloze words. Participants were also more accurate when responding during the Long RTD compared to the Middle and Short RTD, and for the Middle compared to the Short RTD for the high cloze probability words. For the low cloze words, accuracy was higher for the Long RTD compared to both the Middle and Short RTDs, which did not differ from one another. Peak latency of N400 varied as a function of RTD. Specifically, Long RTD elicited earlier peak latencies for the low cloze probability words compared to the Middle and Short RTD. No differences in peak latency were observed between RTDs for the high cloze words. These differences in peak latency suggest that when language demands are greater, such as in the low cloze words, increasing cognitive demand by requiring faster processing for an upcoming response slows processing of the target word. Taken together, the reduced reaction accuracy in conjunction with the slower

N400 latency elicited by the Middle and Short RTDs may reflect the cognitive effects of increasing demand in a listening situation.

Behavior

Reaction Time

No differences in reaction time were observed between the three RTD conditions. However, in the current project, participant reaction times were longer than often reported (typically around 400-800ms). This is likely a result of reaction time also including the time required for participants to read the four visual choices on the monitor prior to responding. There also was not a speed-accuracy trade off, where increased accuracy is associated with slower reaction times, and vice versa, as reaction time did not differ between the RTD conditions. Previous studies with adult with normal hearing (Picou et al., 2011; Desjardins & Doherty, 2013; Downs, 1982) all used a study paradigm that required participants to either repeat target words (Picou et al., 2011; Desjardins & Doherty, 2013) or complete motor tasks while listening to sentences in noise. These studies found significant reaction time delays between easier and harder listening conditions. Studies in which individuals with hearing loss performed tasks of increasing listening difficulty during a dual-task paradigm (working memory, repetition, or other cognitive task) also found that reaction time decreased as subjective LE ratings increased (Rossiter et al., 2006; Mackersie & Cones, 2011; Zekveld et al., 2011). Additionally, elderly populations were found to have decreased reaction times in listening tasks compared to younger populations.

Bertoli & Bodmer (2014) investigated listening effort and the ability to ignore stimuli presented simultaneously with target stimuli, to which older adults with and without hearing loss were asked to press a button. Reaction time between conditions did not differ between the two groups. Bernading et al. (2013) observed some reaction time difference among differing age groups during in a dual-task experiment in which the subjects were asked to repeat a word as quickly as possible while it was presented in noise. With the exception of a few previous studies aforementioned, paradigms consisting primarily of adults with normal hearing found little to no significant differences in reaction time between conditions. Similarly, the current study did not observe differences in reaction time between conditions. This may be a function of the single task paradigm or potentially that small differences between conditions were masked by the increased time required to read the response options.

Performance Accuracy

Participants in the current study more accurately detected words that were highly likely to complete a sentence compared to words that were less likely to finish the sentence. This difference in accuracy as a function of cloze probability has been reported numerous times in the literature (Kutas & Hillyard, 1980). Additionally, the current findings revealed higher reaction accuracy for longer RTD conditions. This finding supports our hypotheses regarding better behavioral performance with decreased cognitive demand, or longer RTDs.

Better reaction accuracy with reduced cognitive demands, or easier listening conditions have been previously reported in the literature. Rackerd et al. (1996) evaluated listening effort in a task in which had participants repeat various spans of digits in noise, during which they found decreased accuracy as the working memory task increased. In contrast, other studies have not observed differences in response accuracy with more difficult listening conditions, despite increased physical responses such as pupil response and skin conductance response

(Krameret al., 1997; Zekveld et al., 2010; Mackersie & Cones, 2011). Kramer and colleagues (1997) evaluated accuracy of story recall in noise and reported decreased accuracy when a higher signal-to-noise ratio that was present. Other studies modulating the difficulty of the listening environment found small differences in performance, in the number of errors produced (Groppe et al., 2010; Sivonen et al., 2006). Benikos et al. (2013) used RTDs similar to the current study in a Go/NoGo task in which the participant would press a button in response to a target shape. As task difficulty increased with shorter RTDs, response accuracy decreased significantly. The current results showing an interaction between RTD and Cloze probability for response accuracy are consistent with these previous findings.

One drawback to the existing literature is that it primarily employed dual-task paradigms to evaluate relationships between performance on a listening task under more or less challenging conditions. Thus, the decrease in accuracy on the secondary task is used to infer increased listening effort (Sarampalis et al., 2009; Downs, 1982; Gosselin & Gagne, 2010; Groupe et al., 2013; Desjardins & Doherty, 2013; Fraser et al., 2010; Hornsby, 2013; Picou et al., 2013). This design potentially limits the degree to which results reflect LE directly. In contrast, the current project used a single-task procedure to evaluate LE under easier and more challenging listening conditions. Together with the existing literature, our current findings suggest that the current paradigm is quantifying LE directly.

N400

Mean Amplitude

Larger N400 mean amplitudes were elicited by the Low Cloze probability than the High Cloze probability sentences. However, no differences were observed in mean amplitudes between the RTD conditions. The N400, which is thought to reflect semantic processing and ease of lexical nitration is also thought to reflect the amount of attention given to processing a word (Kutas & Hillyard, 1980; Kutas & Federmeier, 2000). We predicted that N400 amplitudes would be larger with shorter RTDs, however, our hypothesis was not supported by the data.

Previous LE studies evaluating changes in ERP amplitudes have revealed differences in early exogenous potentials, the N100 and N200 as a function of LE. Increases in N100 amplitudes (Shahin, 2012; Multert et al., 2005) were observed in attend/ignore paradigms in which repetition or recall of sentences in noise was modulated by difficulty. Other studies of endogenous ERP components found few differences in mean amplitude with increased LE (Bernarding et al., 2013; Straus, 2010; Gosselin & Gagne, 2010; Groupe et al., 2013; Pichora-Fuller & Singh, 2006).

In contrast, studies of the later, endogenous ERP potentials, the P300, revealed modulation of ERP amplitudes as a function of LE. P300 amplitudes have been shown to correlate with speech recognition tasks in different listening conditions. Identification of sentences or story elements in varying degrees of noise has revealed decreased P300 amplitudes (Polich & Herbst, 2000; Kramer, et. al, 2005). In contrast, N400 amplitudes elicited by phonemic identification restoration did not reveal any significant N400 amplitude changes as a function of LE (Groppe et al., 2010; Sivonen et al., 2006). While it is difficult to interpret these varying results, our results are consistent with previous findings in N400 amplitudes as a function of LE (Groppe et al., 2010; Sivonen et al., 2006). Although N400 amplitudes increased with shorter RTDs, the difference between RTD conditions was not a significant. These findings indicate that the ease of lexical access or integration does not appear to be significantly impacted by increased cognitive demands in this single-task design.

Peak Latency

N400 peak latency was later for Low cloze compared High cloze sentences, consistent with the existing literature (Kutas & Hillyard, 1980; Kutas & Federmeier, 2000). Longer N400 peak latencies were observed for both the Middle and Short conditions over right hemisphere electrode locations compared to the Long RTD condition. These results support our hypothesis that longer RTDs would require less cognitive effort, and that this would be reflected in the N400 component. However, our hypothesis predicted this would result in differences in N400 mean amplitudes. Instead, we found that conditions requiring less listening effort, specifically the longer RTD, revealed faster lexical access or integration of low cloze probability words. This is reflected by the earlier N400 peak latencies for the Long RTD Low cloze condition. This finding suggests that the increased cognitive load required to respond more quickly in the Middle and Short RTD conditions for the Low cloze condition, which already requires more effortful lexical processing, results in slower neural processing. In other words, increased in cognitive load in more challenging listening conditions delayed language processing time.

Delayed neural responses for early, exogenous ERP components, the N100 and N200, have been observed with increased listening difficulty. This delay is attributed to an immature or decreased processing system which may be present in individuals with hearing loss or age-related variation in cognition (Patel & Azzam, 2005; Pekkonen, Rinne, & Näätänen, 1995; Amenedo, Diaz, 1998; Humes & Wilson, 2003; Humes & Christopherson, 1991). Additionally, delayed peak latencies in later, endogenous components, specifically the P300, have been reported when listeners performed a speech recognition and recall task in a dual-task paradigm in noise (Polich & Herbst, 2000; Kramer, et. al, 2005). Similarly, dual-task studies involving identification of hard-to-hear words/phonemes in noise have reported evaluating longer N400 peak amplitudes for more challenging listening conditions (Groppe et al., 2010; Sivonen et al., 2006; DeLong, Urbach, & Kutas, 2005; Erlbeck et al., 2014).

Consistent with findings in dual-task paradigms, longer N400 latencies were observed with decreased RTD in the current single-task study. These findings suggest that for conditions in which word integration was easier, the high cloze probability condition, for which language processing load is less, it is easier to integrate words even under demanding response conditions with short RTDs. In contrast, when word integration is more difficult and requires more effortful semantic processing, the low cloze probability condition, additional increases on the cognitive demand by reducing the RTD result in slower integration of the final word prior to the response.

LIMITATIONS & FUTURE DIRECTIONS

Several potential limitations must be considered in the current study. The first limitation is the length of the paradigm. Participants were asked to listen to sentences during the ERP portion of the experiment for approximately one to one-and-a-half hours. Decreased motivation or attention may have contributed to reduced accuracy over time. At times, a few participants appeared frustrated or fatigued by the end of the task. Breaks were offered and provided, however, task performance and cognitive processes may have still been affected. The second limitation of this task was that the visual four force choice task increased eye movements and contributed to the large amount of artifact found in the data. Despite images and instructions designed to minimize eye movements, data had to be rejected for eye artifact. Previous studies used a two-forced choice task in conjunction with a second task (Benikos et al., 2013) or a free-choice verbal response (Grouppe et al., 2010; Sivonen et al., 2006) to minimize eye movements, but these studies were dual-task paradigms which resulted in other paradigms, which resulted in other, confounds in the interpretation of their results. The current study used the four forced-choice task to increase response demands on participants in a single-task paradigm. However, future versions of this paradigm might benefit from fewer response options or a different type of response task in order to minimize eye movements while preserving the single-task approach.

Future research can expand and build upon the present procedure. The value of evaluating listening effort by measuring the N400 in a single-task paradigm can be seen directly by the relationship between RTD conditions. As minimal differences were observed between the Middle and Short RTD conditions, perhaps a shorter paradigm with only one longer and one shorter RTD would result in similar N400 results. A shorter paradigm would also make this study design more accessible to other populations (e.g. children, older adults, adults with hearing impairment). The current study also has implications for other populations, including persons with hearing impairment or young children. The paradigm stimuli were developed using words from the MacArthur Communicative Development Inventory (CDI; Fenson et al., 1993) and/or Spoken Word Count frequent words used by five-year-olds (Wepman & Hass, 1969), ensuring that the words in each sentence are child-friendly and would be known to most children and adults. Additionally, the four-talker babble is presented continuously throughout the experiment. Thus, this paradigm can be used to persons with hearing amplification systems that require time to acclimate to background noise.

Using this paradigm in future clinical studies may support the development of a clinically applicable measure of LE. This delayed N400 effect is hypothesized to reflect delayed integration of a word into a sentence when listening conditions are more difficult. The cognitive load associated with increasing the difficulty in both cloze probability and response time deadline was reflected in performance accuracy and slower word response times. There is potential for these findings to be used to develop a non-ERP measure of LE.

CONCLUSIONS

The present study provides initial evidence for the use of a neural index of semantics, the N400, as a marker of listening effort. Differing N400 peak latencies between RTD and cloze probability conditions is associated with increased cognitive requirements for integrating a word into a sentence context. The delayed N400 response is thought to reflect the requirement of increased listening effort for these more challenging listening conditions. Future research is needed to pinpoint the factors that account for the variability observed in neural mechanisms of listening effort.

APPENDIX

APPENDIX A

Table 1. ERP Stimuli

With breakfast, I will drink my orange	**juice**	**shake**
She will get upset if you pull her	**hair**	**ears**
My favorite part about winter is playing in the	**snow**	**sun**
The cow was in the field eating some	**grass**	**paper**
The sprinkler is connected to the garden	**hose**	**bowl**
We go to the chicken house to collect the	**eggs**	**bubbles**
When I asked Jane her age she held up three	**fingers**	**flags**
My favorite is the jelly filled	**donut**	**pancakes**
Man's best friend is a	**dog**	**man**
At the movies, I get	**popcorn**	**cookies**
Before the milk, put the cereal in the	**bowl**	**cup**
Do not forget your umbrella because it might	**rain**	**snow**
Mom could not read without her	**glasses**	**stories**
Please set the table with a knife and	**fork**	**plate**
If it is too bright you can turn off the	**lights**	**TV**
Please, lady, I need you to take your	**seat**	**crib**
After the baseball game, Henry had two black	**eyes**	**cheeks**
For breakfast I just have a single piece of	**toast**	**cake**
Before your shoes, put on your	**socks**	**hat**
At night I look up and see lots of	**stars**	**birds**
At home, I like to talk to my dad and	**mom**	**fish**
After your shower, dry off with the	**towel**	**napkin**
Come over to play at my	**house**	**pool**
I built a sandcastle on the	**beach**	**sidewalk**
To get to school, Jack rode his	**bike**	**sled**
We need to change the baby's	**diaper**	**sweater**
If you want to play catch, throw me the	**ball**	**book**
She placed the bookmark in the	**book**	**box**
To get them clean, wash your hands using the	**soap**	**brush**
When it was raining, my cat watched from the	**window**	**roof**
After spilling the milk, Jess was sent to her	**room**	**beach**
The teacher writes on the board with some	**chalk**	**sauce**
Chuck was very ticklish on the bottom of his	**feet**	**fingers**
When I get hungry between meals I have	**snacks**	**milk**
She answered and said hello into the	**phone**	**banana**

Table 1 cont'd

Monkeys love to eat a	**banana**	**Chocolate**
She laughed because she had broccoli in her	**teeth**	**lunch**
When I am thirsty, I drink	**water**	**pop**
In the fall, we went to the patch to pick a big	**pumpkin**	**rooster**
The farmer road into the field on his	**tractor**	**balloon**
On the walk, the baby sat in the	**stroller**	**tractor**
Instead of getting in the bath, I want to get in the	**shower**	**washer**
Please throw your wrapper the	**trash**	**purse**
For our picnic, we laid out the	**blanket**	**doll**
Popsicles can be messy so use a	**napkin**	**hose**
There are no more cookies left in the	**jar**	**zoo**
Saturday is the best day to see the animals at the	**zoo**	**farm**
Around the campfire, we trade	**stories**	**snacks**
The butterfly landed on all the pretty	**flowers**	**pillows**
My bunny wants to eat	**carrots**	**cheese**
At recess, we pushed each other on the	**swings**	**car**
Joe watched TV sitting on the	**couch**	**swings**
At night, she puts the baby in the	**crib**	**seat**
I will only eat toast with lots of grape	**jelly**	**juice**
The girl loves to play outside and blow	**bubbles**	**beads**
We picked vegetables from the neighbors'	**garden**	**trash**
Flapping in the wind, were several American	**flags**	**ducks**
On the lake, I get to go out in the	**boat**	**woods**
To eat my soup, I opened the	**microwave**	**fridge**
Julie loves to put syrup on her	**pancakes**	**bread**
We are having the picnic in the	**park**	**dark**
When my hair looks bad I wear a	**hat**	**Scarf**
You need sunscreen if you are playing in the	**sun**	**tree**
The deer ran to his home in the	**woods**	**yard**
Please hand me a pencil and some	**paper**	**flowers**
Help me make hamburgers for	**dinner**	**mom**
Her bed was covered in several fluffy	**pillows**	**clouds**
They made a necklace by stringing together some	**beads**	**beans**
I saw a squirrel climbing in the	**tree**	**stroller**
I hung my jacket up in the	**closet**	**garden**
Please get the milk from the	**fridge**	**table**
The zebra ran from the	**lion**	**elephants**

Table 1 cont'd

If there is no flavor, add a little	**salt**	**jelly**
I need to go get eggs from the	**store**	**school**
Elephants have really large	**ears**	**teeth**
I cannot wait to see the cows on the	**farm**	**ball**
We should watch a movie on the	**TV**	**phone**
I leaned over and kissed her on the	**cheeks**	**legs**
The best thing about summer is cooling off in the	**pool**	**bathroom**
Looking into the water, I saw huge	**fish**	**stars**
For the bake sale, I will bring	**cookies**	**candy**
Every night we have dinner at the	**table**	**park**
We gave the snowman a colorful	**scarf**	**tail**
In the backyard collecting nuts was a	**squirrel**	**pig**
Rolling in the mud on the farm was a	**pig**	**sheep**
Please get her wallet of her	**purse**	**closet**
It is cold outside and I forgot my	**jacket**	**pants**
Racing to the fire, we got out of the way for the	**fire truck**	**train**
The lunch was burger, fries and	**pop**	**ice cream**
It is almost time for them to cut the	**cake**	**hair**
I need to get the kids to	**school**	**dinner**
His favorite ball rolled into the	**street**	**store**
When they moved, they put their stuff in a	**box**	**towel**
On top of spaghetti we put lots of	**sauce**	**eggs**
I want to go to the lake to feed the	**ducks**	**frogs**
For my garden, my husband gave me some rose	**bushes**	**soap**
Her favorite meal is to mix rice and	**beans**	**pumpkin**
You will find the tissue in the	**bathroom**	**grass**
On a hot day, Sam went to get	**ice cream**	**popcorn**
In the winter, we ride on the	**sled**	**fire truck**
The horse has a long brown	**tail**	**blanket**
After running, Maggie had very wobbly	**legs**	**eyes**
I fell and ripped my	**pants**	**socks**
What you really need for sandwiches is	**bread**	**toast**
She drew with chalk on the	**sidewalk**	**porch**
If your hair is messy, you could use a	**brush**	**shower**
The rain is dripping off of the	**roof**	**chair**
After a hard day, I could use a big	**hug**	**bed**
Meet me at noon to get your	**lunch**	**bike**
She put the cookies for the party on a	**tray**	**jar**

Table 1 cont'd

I used the sink to wash my dirty	dishes	feet
She is over there by the woman and the	man	coffee
The little girl had a tea party with her	doll	dog
In my party bag, I got a bunch of	candy	carrots
It is so nice out, let's play in the	yard	Street
So food does not stick to the pan, add	butter	salt
The kids sounded like a herd of	elephants	deer
I like to read before I get in the	bed	boat
Getting ready for dinner, he pulled up a	chair	window
When the baseball hit his face, the boy broke his	nose	chalk
Luna's shirt got caught in her	zipper	diaper
The cheese is set in the trap for the	mouse	house
You will not get in trouble if you do it in the	dark	light
In the woods, there might be a	deer	jacket
At the party, I let go of the string that held my red	balloon	nose
We watched the sunset from our	porch	couch
My dessert needs to be cake that tastes like	chocolate	water
The sheppard and the farm dog gathered up the	sheep	sticks
Please put some coffee in my	cup	kitchen
She got a new hat and matching	gloves	zipper
Empty your dirty clothes hamper into the	washer	bushes
We put up a scarecrow in the garden to scare the	birds	mouse
I went to the bakery to get a donut and a	coffee	hug
On the farm, we get woken up by a loud	rooster	squirrel
Please help me pull on my wool	sweater	gloves
My favorite part about flying is seeing all the	clouds	rain
I like macaroni and	cheese	butter
After dinner, I would like to get a large chocolate	shake	donut
We roasted the marshmallows using a long	stick	fork
I like to dip my cookies in my	milk	room
I sat next to the tracks and saw a big	train	lion
If we want to drive, we need to get my	car	glasses
By the pond, we found the green	frogs	dishes
Dad hit the nail in the wall using a	hammer	tray
Mom got the cookies out of the	oven	pajamas
Jane put the pieces of pizza on the	plates	dryer
To warm up my slippers, I put them in the	dryer	oven
In Mom's purse you'll find her bright, red	lipstick	hammer
At the sleepover, the girls put on their	pajamas	lipstick

Table 1 cont'd

After making dinner, I cleaned the whole	**kitchen**	**microwave**

APPENDIX B

Table 2. Calibration Stimuli with Low Cloze Ending Only

Hopping through the field was a small gray	**cat**
I traveled to grandma's house riding on a	**motorcycle**
To finish painting the sign, I need more	**glue**
Every morning for breakfast, I have a bowl of	**nuts**
With my hamburger, I would like a side of	**noodles**
For lunch, we ordered a delivery of	**chicken**
Before going out in the rain, Haley put on her rain	**shorts**
The best part about going to the party was wearing my new	**slippers**
In the woods, we heard a howling	**owl**
When we went to the farm, we rode on a	**donkey**
When Jane cried, she put her head between her	**arms**
Karen put red makeup around her	**face**
Greg fell during soccer and scraped his	**nose**
Mary cleaned the kitchen floor with a	**toothbrush**
She lives in the house across the	**backyard**
The cat ran through the house to catch the	**ant**
I set my morning alarm on my	**stove**
When I went to the beach, I collected colorful	**rocks**
To check the weather, I turn on the	**radio**
Maggie tripped and fell going up the	**ladder**
We usually eat dinner sitting at the	**sink**
To get the car, she needs to have more	**people**
Harry cut the wood with sharp	**scissors**
At school, I took a test given by the	**fireman**
When Cam got sick, he went to his	**brother**
Many animals go out and play when they see the	**moon**
Little boys love to eat many	**apples**
The clown turned the balloons into a	**turtle**
The little girl could not sleep without her stuffed	**turkey**
To hang the picture on the wall you need some	**tape**

REFERENCES

REFERENCES

Allison, B. Z., & Polich, J. (2008). Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. *Biological psychology*, *77*(3), 277-283.

Amenedo, E., & Dıaz, F. (1998). Aging-related changes in processing of non-target and target stimuli during an auditory oddball task. *Biological Psychology*, *48*(3), 235-267.

ANSI S3.6 (2010). Specification for Audiometers. American National Standards Institute.

Beck, DL. & Flexer C. (2011). Listening is where hearing meets brain: In children and adults. *Hearing Review*, February: 30-33.

Benikos, N., Johnstone, S., & Roodenrys, S. (2013). Varying task difficulty in the Go/Nogo task: the effects of inhibitory control, arousal, and perceived effort on ERP components. *International Journal of Psychophysiology, 87,* 262-272.

Bernarding, C., Strauss, D., Hannenmann, R., Seidler, H., & Corona-Strauss, F. (2013). Neural correlates of listening effort related factors: Influence of age and hearing impairment. *Brain Research Bulletin, 91,* 21-30.

Bertoli, S. & Bodmer, D. (2014). Novel sounds as a psychophysiological measure of listening effort in older listeners with and without hearing loss. *Clinical Neurophysiology, 125,* 1030-1041.

Brown, L., Sherbenou, R., & Dollar, S. (1982). Test of nonverbal intelligence: 3rd edition (TONI-3). PRO-ED, Austin, TX.

Dawes, P., Munro, K. J., Kalluri, S., & Edwards, B. (2014). Acclimatization to hearing aids. *Ear and hearing*, *35*(2), 203-212.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature neuroscience*, *8*(8), 1117-1121.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods 134, 9–21.

Desjardins, J. L., & Doherty, K. A. (2013). Age-related changes in listening effort for various types of masker noises. *Ear and hearing*, *34*(3), 261-272.

Donchin, E., & Coles, M. G. (1998). Context updating and the P300. *Behavioral and brain sciences*, *21*, 152-154.

Downs, DW. (1982). Effects of hearing and use on speech discrimination and listening effort. *Journal of Speech and Hearing Disorders, 47,* 189–193.

Downs, DW., Crum, MA. (1978). Processing demands during auditory learning under degraded listening conditions. *Journal of Speech and Hearing Research, 21,* 702–714.

Erlbeck, H., Kubler, A., Kotchoubey, B., & Veser, S. (2014). Task instructions modulate the attentional mode affecting the auditory MMN and semantic N400. *Frontiers in Human Research, 8,* 2-16.

Fraser, S., Gagne, J. P., Alepins, M., & Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: The effects of providing visual speech cues. *Journal of Speech, Language, and Hearing Research*, *53*(1), 18-33.

Fournier, L. R., Wilson, G. F., & Swain, C. R. (1999). Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training. *International Journal of Psychophysiology*, *31*(2), 129-145.

Fenson, L., Dale, P.S., Reznick, J.S., Thal, D., Bates, E., Hartung, J.P., Pethick, S., & Reilly, J.S. (1993). The MacArthur Communicative Development Inventories: User's guide and technical manual. San Diego, CA: Singular Publishing Group.

Feuerstein, J. F. (1992). Monaural versus binaural hearing: ease of listening, word recognition, and attentional effort. *Ear and Hearing*, *13*(2), 80-86.

Gatehouse, S. & William, N. (2004). The speech, spatial, and qualities of hearing scale (SSQ). *International Journal of Audiology, 43,* 85-99.

Golden, C., Freshwater, S., & Golden, Z. (2003). Stroop color and word test. Stoelting Co. Wood Dale, IL.

Gosselin, P. & Gagne, P. (2010). Use of dual-task paradigm to measure listening effort. *Canadian Journal of Speech-Language Pathology and Audiology, 34,* 43-51.

Groppe, D. M., Choi, M., Huang, T., Schilz, J., Topkins, B., Urbach, T. P., & Kutas, M. (2010). The phonemic restoration effect reveals pre-N400 effect of supportive sentence context in speech perception. *Brain research*, *1361*, 54-66.

Hagoort, P. & Brown, C. (2000). ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia, 38,* 1518-1530.

Hällgren, M., Larsby, B., Lyxell, B., & Arlinger, S. (2001). Evaluation of a cognitive test battery in young and elderly normal-hearing and hearing-impaired persons. *Journal of the American Academy of Audiology*, *12*(7), 357-370.

Hammill, D.D., Brown, V.L., Larsen, S.C., Wiederholt, J.L. (1994). Test of adolescents and adult language-third edition (TOAL-3). Pearson, New York, NY.

Harmony, T., Bernal, J., Fernández, T., Silva-Pereyra, J., Fernandez-Bouzas, A., Marosi, E., & Reyes, A. (2000). Primary task demands modulate P3a amplitude. *Cognitive Brain Research*, *9*(1), 53-60.

Hicks, C. B., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *Journal of Speech, Language, and Hearing Research*, *45*(3), 573-584.

Hornsby, B. W. (2013). The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear and Hearing*, *34*(5), 523-534.

Kiessling, J., Pichora-Fuller, K., Gatehouse, S., Stephens, D., Arlinger, S., Chisolm, T., et al. (2003). Candidature for and delivery of audiological services: special needs of older people. *International Journal of Audiology*, 42, 2, S92-S101.

Koelewijn, T., de Kluiver, H., Shinn-Cunningham, B. G., Zekveld, A. A., & Kramer, S. E. (2015). The pupil response reveals increased listening effort when it is difficult to focus attention. *Hearing research*, *323*, 81-90.

Krause, M., Kennedy, M., & Nelson, P. (2014). Masking release, processing speed and listening effort in adults with traumatic brain injury. *Brain Injury, 28,* 1473-1483.

Kramer, S. E., Kapteyn, T. S., Festen, J. M., & Kuik, D. J. (1997). Assessing aspects of auditory handicap by means of pupil dilatation. *International Journal of Audiology*, *36*(3), 155-164.

Kramer, A.F., Trejo, L.J., Humphrey, D., 1995. Assessment of mental workload with task-irrelevant auditory probes. Biological Psychology 40, 83–100.

Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, *50*(1), 23-34.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207,* 203-205.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in cognitive sciences*, *4*(12), 463-470.

Kutas, M., & Federmeier, K. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology, 62,* 621-647.

Larsby, B., Hällgren, M., Lyxell, B., & Arlinger, S. (2005). Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *International Journal of Audiology*, *44*(3), 131-143.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920-933.

Lopez-Calderon, J., Luck, S., 2010. ERPLAB Toolbox (1.1.0). Retrieved from http://erpinfo.org/erplab

Ma, Q., Meng, L., Wang, L., & Shen, Q. (2014). I endeavor to make it: Effort increases valuation of subsequent monetary reward. *Behavioural Brain Research, 261*, 1-7.

Mackersie, C. L., & Cones, H. (2011). Subjective and psychophysiological indices of listening effort in a competing-talker task. *Journal of the American Academy of Audiology*, *22*(2), 113.

Matsuda, I., & Nittono, H. (2014). Motivational significance and cognitive effort elicit different late positive potentials. *Clinical Neurophysiology*.

Mulert, C., Menzinger, E., Leicht, G., Pogarell, O., & Hegerl, U. (2005). Evidence for a close relationship between conscious effort and anterior cingulate cortex activity. *International Journal of Psychophysiology*, *56*(1), 65-80.

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.

Pekkonen, E., Rinne, T., & Näätänen, R. (1995). Variability and replicability of the mismatch negativity. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *96*(6), 546-554.

Pichora-Fuller, M., & Singh, G. (2006). Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiological rehabilitation. *Trends in Hearing, 10,* 29-89.

Picou, E. M., Ricketts, T. A., & Hornsby, B. W. (2013). How hearing aids, background noise, and visual cues influence objective listening effort. *Ear and hearing*, *34*(5), e52-e64.

Polich, J., & Herbst, K. L. (2000). P300 as a clinical assay: rationale, evaluation, and findings. *International Journal of Psychophysiology*, *38*(1), 3-19

Pritchard, W. S., Shappell, S. A., & Brandt, M. E. (1991). Psychophysiology of N200/N400: A review and classification scheme. *Advances in psychophysiology*, *4*, 43-106.

Rakerd, B., Seitz, P., & Whearty, M. (1996). Assessing the cognitive demands of speech listening for people with hearing losses. *Ear and Hearing*, *17*(2), 97-106.

Rossiter, S., Stevens, C., & Walker, G. (2006). Tinnitus and its effect on working memory and attention. *Journal of Speech, Language, and Hearing Research*, *49*(1), 150-160.

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*, *52*(5), 1230-1240.

Shahin, A. J., Kerlin, J. R., Bhat, J., & Miller, L. M. (2012). Neural restoration of degraded audiovisual speech. *Neuroimage*, *60*(1), 530-538.

Sivonen, P., Maess, B., Lattner, S., & Friederici, A. D. (2006). Phonemic restoration in a sentence context: evidence from early and late ERP effects. *Brain research*, *1121*(1), 177-189.

Strauss, D. J., Corona-Strauss, F. I., Trenado, C., Bernarding, C., Reith, W., Latzel, M., & Froehlich, M. (2010). Electrophysiological correlates of listening effort: neurodynamical modeling and measurement. *Cognitive neurodynamics*, *4*(2), 119-131.

Ullsperger, P., Freude, G., Erdmann, U., 2001. Auditory probe sensitivity to mental workload changes – an event-related potential study. *International Journal of Psychophysiology 40,* 201–209.

van de Meerendonk, N., Kolk, H. H., Vissers, C. T. W., & Chwilla, D. J. (2010). Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, *22*(1), 67-82.

Weisz, N., & Obleser, J. (2014). Synchronisation signatures in the listening brain: a perspective from non-invasive neuroelectrophysiology. *Hearing research*, *307*, 16-28.

Wepman, J. M., & Hass, W. (1969). A Spoken Word Count (Children--Ages 5, 6 and 7).

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation. *Ear & Hearing*, *20*, 00-00.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and hearing*, *31*(4), 480-490.