Audiovisual Spoken Word Processing in Typical-Hearing and Cochlear Implant-Using Children: an ERP Investigation

By

ELIZABETH PIEROTTI
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

David P. Corina, Chair

Katharine Graf Estes

Tamara Swaab

Committee in Charge

2024

Acknowledgements

I want to express my appreciation for my advisor, Dr. David P. Corina, for his guidance and enthusiasm throughout this research. His support on this project has been invaluable. I'm extremely thankful to have learned from him during my graduate training and I couldn't have produced research of this caliber without his direction.

Many others in the Cognitive Neurolinguistics Lab contributed to this research. Most notably is Sharon Coffey-Corina, without whom this study could not have happened. I'm grateful to have learned how to collect, analyze and deeply understand EEG/ERP data from a true master. My lead research assistant, Sana Shehabi, played a crucial role in data collection, analysis, and providing feedback to this research. I am thankful for her help and excited to "pass the torch" onto her. Tristan Schaefer was instrumental in the development of the experiment, and many others gave feedback on early iterations of the study and the preliminary findings. Thank you all so much for your contributions.

Endless thanks to my committee members Dr. Katharine Graf Estes and Dr. Tamara Swaab for their patience and encouragement during this process, and for the time they have invested in reviewing and improving this work; to Dr. Lisa Oakes and Dr. Ron Mangun for the pivotal roles they have played in my professional development outside of the lab; to Angela Scully for her commitment to my success and well-being throughout my graduate experience; and to fellow grad student Alexandra Theodorou for her camaraderie as we have navigated graduate school together.

Thank you to those outside of grad school who have uplifted me over the last six years. My mom and dad were very kind and supportive when I moved to California to attend UC Davis, and their encouragement has never wavered. Thank you to my sisters, Caroline and Stephanie; my grandparents; my friends, especially Annie Matthews; and my dog, Pepperoni. I would not have made it this far without you!

Abstract

The process of spoken word recognition is influenced by both bottom-up sensory information and top-down cognitive information. These cues are used to process the phonological and semantic representations of speech. Several studies have used EEG/ERPs to study the neural mechanisms of children's spoken word recognition, but less is known about the role of visual speech information (facial and lip cues) on this process. It is also unclear if populations with different early sensory experiences (e.g. deaf children who receive cochlear implants; CIs) show the same pattern of neural responses during audiovisual (AV) spoken word recognition. Here we investigate ERP components corresponding to typical hearing (TH) and CI-using school age children's sensory, phonological, and semantic neural responses during a picture-audiovisual word matching task. Children (TH n = 22; CI n = 13; ages 8 – 13 years) were asked to match picture primes and AV video targets of speakers naming the pictures. ERPs were time-locked to the onset of the target's meaningful visual and auditory speech information. The results suggest that while CI and TH children may not differ in their sensory (Visual P1, Auditory N1) or semantic (N400, Late N400) responses, there may be differences in the intermediary components associated with either phonological or strategic processing. Specifically, we find an N280 response for the CI group and a P300 component in the TH group. Subjects' ERPs are correlated with their age, hearing experience, task performance, and language measures. We interpret these findings in light of the unique strategies that may be employed by these two groups of children based on the utilization of different speech cues or task-level predictions. These findings better inform our understanding of the neural bases of AV speech processing in children, specifically where differences may emerge between groups of children with differential sensory experiences; the results have implications for improving spoken language access for children with cochlear implants.

Contents

List of Tables and Figures

Tables

Figures

Introduction:

Severe to profound hearing loss that occurs at birth or early in development can pose challenges for children's language development. Deaf children with deaf or hearing parents who use sign languages have access to a rich linguistic environment early on, but over 90% of deaf children are born to hearing parents who do not sign (Lederberg et al., 2013; Mercure et al., 2019). Consequently, deaf children with parents who do not sign face the possibility of experiencing delays in early language experience. Without sufficient linguistic input (signed and/or spoken) in the first few years of life, children are at risk for atypical development of both first and subsequent languages (Mayberry, 2007), reading skills (Geers & Hayes, 2011), and socio-cognitive processes such as theory of mind (Meristo et al., 2016; Peterson et al., 2005). Difficulty in the use of these fundamental skills can lead to downstream challenges in educational and social achievement. This situation has been termed "language deprivation syndrome" (Hall, 2017). In order to avoid outcomes associated with this, it has become apparent that a fully-accessible first language foundation (either signed or spoken) is imperative for congenitally deaf children to be able to experience healthy growth in all developmental domains.

Technological advances have led to interventions for early childhood hearing loss. Cochlear implants (CIs) are an increasingly popular option by which deaf and hard-of-hearing individuals can gain access to spoken language, as they generally offer better auditory input (e.g., a wider range of frequencies) for the development of oral/spoken language as compared to hearing aids (Connor et al., 2006; Niparko et al., 2010). CIs directly stimulate the auditory nerve through electrical impulses (Moore & Shannon, 2009). Incoming acoustic information is processed by the CI which encodes the speech envelope (composed of temporally delimited complex spectral-frequency signals) to a simulation pattern that approximates the spectral properties of the acoustic event. More specifically, following the tonotopic organization of a typical basilar membrane within the cochlea, an electrode is placed along the basilar membrane with localized channels that can stimulate different areas of the basilar membrane in response to different perceived frequencies. This allows for the perception of complex combinations of frequencies, such as those found in speech signals. Despite some distortion and limitations in bandwidth, CIs can be effective at transmitting auditory information (McMurray et al., 2017).

Congenitally deaf children who receive CIs, however, can have variable success in speech processing, even with years of experience using their device (Dunn et al., 2014; Geers et al., 2011; Schorr et al., 2008; Wie et al., 2007). Wie and colleagues (2007) followed a cohort of 79 CI-using deaf children several years after receiving their implant and found a wide range of speech recognition capabilities, even up to five years after implantation (Wie et al., 2007). Children who receive their implant early in development and who have extended use with their device tend to show better spoken language outcomes than other CI-using children (Dunn et al., 2014; Geers et al., 2011; Yoshinaga-Itano et al., 2010). Considering vast individual differences in etiology, linguistic experience, and device specifications, it may seem daunting to identify the source (or sources) of this variability amongst CI users. In order to gain insight into this complex issue, we presently focus on the spoken word recognition process and seek to understand potential differences at a neural level amongst CI-using children and typically-hearing (TH) children.

Foundational Models of Spoken Word Recognition

Historically, models of spoken word recognition argue for an initial processing of incoming auditory information that generates an internal representation computed from this auditory input. The nature of this representation has been contested, as some have proposed that it may consist of temporally-defined spectral templates, phonemes, or syllables (Frauenfelder & Tyler, 1987; Mehler et al., 1981; Pisoni & Luce, 1987). Regardless of the nature of this representation, once it has been generated, it can make contact with lexical word-level knowledge. Theoretically, the richer the nature of this contact representation, the fewer number of lexical entries are contacted given a more descriptive and discriminative input (Frauenfelder & Tyler, 1987). This contact finally provides the opportunity to arrive at the meaning of a word.

In this sense, inputs are provided to the spoken word recognition system in a linear fashion and speech is recognized as it unfolds. According to the Cohort model, the number of potential word candidates becomes increasingly constrained as the phonological information in a word is processed, until only a single candidate remains (Marslen-Wilson & Tyler, 1980). The Shortlist model (Norris et al., 2000) also argues for the importance of feed-forward or bottom-up connections. This model argues that a "short list" of word candidates is designated from lexical entries that share phonological features with the speech input. Next these word candidates compete and inhibit one another at a lexical level of processing in order to be recognized or selected.

The bottom-up word recognition mechanisms as described by the Cohort and Shortlist models, however, only account for some of the instances of competition that can occur during phonological processing. Pre-lexical phonological competition, such as cohort effects, can be influenced by top-down contextual constraints. Some of these top-down mechanisms are described by models of continuous spoken word recognition such as the Neighborhood Activation Model (NAM) and TRACE. As outlined by Luce and Pisoni (1998), NAM predicts that words must be recognized in the context of other words, with specific emphasis on the frequency and similarity amongst words (e.g., neighborhood density). These factors can influence the speed and accuracy with which one can recognize a given word. While NAM can account for broad types of phonological competition as with rhymes, it does not account for temporal aspects of the spoken word recognition process in a way that models such as TRACE can.

The TRACE model as proposed by McClelland and Elman (1986) argues for the temporal and dynamic nature of spoken language such that both pre-lexical and lexical mechanisms interact to guide recognition. In other words, competition can occur amongst candidates within a given layer at any point during the spoken word recognition process, whether before or after lexical access. Pre-lexical forms of competition may occur amongst items that have overlapping speech sounds (e.g., rhymes or word-initial cohorts) but competition can also occur based on lexical knowledge such as word frequency. Additionally, connections exist between layers of the speech processing hierarchy. Consider how lexical knowledge of word frequency can influence the identification of phonemes: “ball” occurs much more frequently in the English language than “pall”, and when the phonological onset of the word is ambiguous, we are much more likely to assume that the word that was presented was the word with a higher frequency. In this sense, many types of lexical competition can occur during the recognition of a spoken word under the TRACE model (Joanisse & McClelland, 2015; McClelland & Elman, 1986).

The Role of Prediction in Spoken Word Recognition

A prominent theme in the theories described above is the concept of "activation," such that within a network, lexical candidates (words) correspond to individual nodes (McClelland & Elman, 1986; Norris, 1994). These nodes have an assigned activation value that can increase as more perceptual input is received and decrease due to inhibition from other words. Some have argued, however, that activation has little direct correspondence with behavioral observations such as accuracy, probability, or speed of word recognition/response. The concept of activation may be used in computations within the above theories to help estimate responses, but ultimately activation as a metaphor may overcomplicate our interpretation of how a model relates to behavior (Norris & McQueen, 2008).

Some recent models of spoken word recognition have favored Bayesian principles over the concept of activation. Bayesian inference offers a path to optimal spoken word recognition in the face of ambiguous perceptual input by combining prior knowledge of the probabilities of words with available perceptual evidence (Norris & McQueen, 2008; Weber & Scharenborg, 2012). These models rely on Bayesian inference for further explanation of effects such as word frequency by arguing that frequency is derived from one's knowledge of prior probabilities.

Pulling from the action perception literature, Pickering and Garrod (2013) argue for a unified model of language production and comprehension that draws heavily upon one's experience both comprehending and producing language in order to form predictions about incoming spoken language input. Predictions about linguistic input may be made at different levels in the speech processing hierarchy (e.g., phonological, semantic, syntactic). Similar to action perception, listeners may simulate what they themselves would say in a given situation in order to form a prediction about what another speaker is going to say, using a forward model. Therefore, spoken word recognition and production are inherently linked, and context and the ability to represent what another speaker might say is imperative in the successful prediction of spoken input (Pickering & Garrod, 2013).

A predictive coding model of single spoken word recognition has emerged as an alternative to models that rely on activation and inhibition amongst lexical candidates in order to identify a spoken word. Broadly speaking, predictive coding theory asserts that the brain uses top-down processing to anticipate current sensory input and influence lower levels of processing (Engel et al., 2001; Gagnepain et al., 2012; Maess et al., 2016; Rauss & Pourtois, 2013). In the context of speech recognition, Gagnepain and colleagues (2012) posit that lexical candidates that match incoming speech signals generate predictions for upcoming speech segments, which are either confirmed or disconfirmed. If the incoming sensory input matches expectations, processing demands at lower levels of the speech processing hierarchy will be reduced. If predictions are incorrect, however, speech recognition performance will be negatively affected, given that some actions based on the predicted input may have already been initiated. When predictions are weak or absent (e.g., in the face of poor perceptual input or limited context), speech recognition performance may not be affected.

Others have suggested that under certain circumstances, information at lower levels of representations may be *predictively pre-activated* by higher-level inferences, before incoming bottom-up information reaches these lower levels (Kuperberg & Jaeger, 2016). This concept shares similarity with the feed-forward and feed-back levels of influence that are discussed in connectionist models like TRACE. Predictive pre-activation, however, might be a costly strategy that is only used when there is high uncertainty about the bottom-up input, such as in a noisy environment. Importantly, prediction in the context of spoken language comprehension is argued not to be an all-or-nothing phenomenon, but instead is utilized in a graded nature dependent on multiple factors (e.g., contextual constraint, probabilistic knowledge, and the fidelity of bottom-up input; Brothers et al., 2019).

Neurocognitive Approaches to Spoken Word Recognition

The previously summarized recent models of spoken word recognition all support the concept of prediction, in some form, during the process of comprehending spoken words. What is less clear is the neural source of predictive coding. Several studies (Gagnepain et al., 2012; Maess et al., 2016) use evidence from magnetoencephalography (MEG) to propose that the difference between lexical predictions and the current speech input (the "prediction error") is coded in the superior temporal gyrus (STG). STG cells then project the prediction error to higher levels in the speech processing hierarchy in order to update previously compatible lexical-semantic representations.

Hickok and Poeppel (2007) proposed a comprehensive dual stream model for speech processing that mirrors a similar concept to the widely-cited dual stream model of visual perception. In this model, early speech processing occurs in bilateral auditory regions in the dorsal STG (spectrotemporal analysis) and the mid-posterior temporal sulcus (STS; phonological network), before splitting into two separate pathways that support different functions of speech processing. The dorsal pathway is implicated in the motor-mapping of incoming speech (premotor cortex) and multisensory processing (posterior STS; Beauchamp, 2016). The ventral pathway, on the other hand, is more often associated with meaning-mapping of incoming speech, such that this pathway would lead to cortical areas associated with semantic memory and comprehension (such as the middle temporal gyrus and the inferior temporal sulcus, MTG and ITS respectively; Hickok, 2012, 2022). Importantly, this model suggests that the neural organization of speech processing is task-dependent, such that “speech perception” tasks (those that involve sublexical processes such as syllable discrimination) may map to different neural areas than “speech recognition” tasks (tasks that occur after the transformed acoustic representation makes contact with the lexical-semantic level).

In addition to the structures and systems implicated by the dual stream model above, neuronal rhythms that structure the electrical activity in the brain have garnered attention for their potential causal link to speech processing. Similar to the speech processing hierarchy, oscillations in the brain can be categorized in a hierarchical fashion (Poeppel & Assaneo, 2020). Specifically, 30-80 Hz gamma oscillations are associated with phonetic analysis, 4-9 Hz theta oscillations regulate syllabic segmentation, and 1-2 Hz delta oscillations are associated with processing prosodic and syntactic information (Nabé et al., 2021). In this sense, lower frequencies of oscillations are linked to higher levels within the speech processing hierarchy. An exception to this is top-down predictive processing, which may occur at many different levels within the theoretical speech processing hierarchy, and has been suggested to be connected to the beta band (15-20 Hz) as a channel for predictions to descend to lower levels of processing. Neural oscillations are integral to the predictive coding framework detailed above (Engel et al., 2001; Gagnepain et al., 2012; Rauss & Pourtois, 2013), in that oscillations may provide the neural mechanism by which top-down predictions are integrated with bottom-up information.

Visual Speech Cues

The above models of spoken word recognition cover a wide array of behavioral phenomena and provide several compelling options for neural mechanisms that may drive the patterns of behavior observed. None of these models, however, account for the contextual and social cues that can aid (or impair) a listener as they comprehend naturalistic spoken language. Speech processing is affected by bottom-up social information such as dialect and speaking style, as well as top-down contextual information such as speaker familiarity and semantic predictability (Dossey et al., 2023). Scott (2019) provides an important reminder that "speech is an overwhelmingly social behavior" (pg. 58), and that flexibility in utilization of various speech cues is imperative as listeners are constantly exposed to new speakers/dialects, complex acoustic listening environments, and various emotional- and identity-dependent semantic content - all of which may influence speech processing.

More often than not, speech provides both audio *and* visual information, such that facial and lip cues are available along with auditory speech input. There is abundant behavioral evidence that audiovisual (AV) speech is easier to recognize than auditory-only (AO) speech, by way of combining incoming sensory inputs through multisensory integration. This is especially true in noisy environments, when the speaker's message is ambiguous, or when listening to speech that is not of one's native language (Barutchu et al., 2010; Bernstein et al., 2004; Navarra & Soto-Faraco, 2005; Reisberg et al., 1987; Ross et al., 2007). In these situations, auditory input on its own may be less reliable or readily available to the listener; therefore, visual cues may be of use in speech processing. Of relevance to the present research, AV speech can facilitate recognition in populations for whom the processing of auditory information may differ from that of typically developing children, specifically congenitally deaf children who make use of cochlear implants to comprehend spoken language (Geers et al., 2003; Lachs et al., 2001).

Given the wealth of behavioral and neurophysiological (see below) evidence supporting the utility of visual speech cues, it is necessary to reconcile the role of visual information with pre-existing and more current models of spoken word recognition. The question remains: at what point during the spoken word recognition process is visual speech information used? Visual cues may influence the initial sensory processing of auditory signal (Van Wassenhove et al., 2005), the parsing of speech sounds during phonological processing (Baart & Samuel, 2015), or during the access of word forms during lexical-semantic processing (Brunellière et al., 2020; Fort et al., 2013; Mattys et al., 2002).

Congruent visual speech information has been shown to speed up the processing of incoming auditory information (Alsius et al., 2014; Besle et al., 2004; Van Wassenhove et al., 2005), possibly by contributing to the constraints placed on the sensory processing system by internal predictions (top-down processing). Visual cues can often even appear prior to the onset of auditory signal by 10s to 100s of milliseconds. This is highly variable based on the particular word being produced, however, such that words beginning with /m/ speech sounds can have auditory onset prior to the appearance of any visual articulatory movements while words beginning with /h/ sounds may have visual cues that appear over 100 ms prior to the auditory onset of the word. This relationship between visual and auditory speech cue onset is also dependent on the phoneme's position within the utterance (Schwartz & Savariaux, 2014).

It is important to note that there is some consistency in the movements that speakers provide during the production of specific phonemes. The term *viseme* was coined by Fisher (1968) to serve as a visual counterpart to a phoneme, such that visemes are the smallest unit of visual speech information that can differ based on the speech sound being produced (e.g., /p/ and /b/ both share a bilabial place of articulation and have more similar visemes to each other than /f/ or /s/). Visemes can differ in the number of speech sounds they represent, leading to certain visemes being less informative than others. In other words, there is more ambiguity or confusion associated with visemes that correspond to many speech sounds (Fisher, 1968). Following this logic, visual speech information in the form of visemes might modulate prelexical phonological processing, such that the activation and inhibition of particular visual speech representations can influence the subsequent activation of lexical word-form candidates. This, however, may be dependent upon the viseme itself, in that this influence may only occur when visual information is available before or at the same time as acoustic information. More concretely, following the example that was given above, words that begin with the /h/ sound may be more susceptible to modulation from visual information because visemes are available prior to the onset of auditory information, relative to words that begin with /m/ sounds where visual cues appear after auditory information.

Interactions between lip-reading and lexical characteristics such as neighborhood density and frequency have been reported (Fort et al., 2013; Mattys et al., 2002). In tasks using both visual-only and audiovisual speech (Mattys et al., 2002; Tye-Murray et al., 2007), words were identified more accurately when they had fewer visual articulatory competitors; in short, it was easier for participants to lipread words that were comprised of more distinct visemes or did not share visual features with many other words. Additionally, visual information was found to facilitate recognition of low-frequency words more than high-frequency words, suggesting that visual speech cues may be particularly helpful in situations where lexical access is difficult (Fort et al., 2013).

In sum, there is evidence that implicates the use of visual speech information at all stages of spoken word recognition. However, Baart and Samuel (2015) argue that speech processing benefits provided by lip-reading are dissociable from the benefits provided from lexical-semantic context, such that that visual speech cues may facilitate phonological processing but may not influence lexical access. This claim is supported by both behavioral and electrophysiological evidence (Baart et al., 2014; Baart & Samuel, 2015), and has been reconciled with the dual stream model of speech processing proposed by Hickok & Poeppel (2007). Following the dual-stream model, it is plausible that context provided by visual speech cues would be processed in a separate stream (the dorsal stream, associated with multisensory processing) from the processes that occur at a lexical-semantic level in the ventral stream.

The present study contributes to the limited research on spoken word recognition that uses naturalistic audiovisual speech stimuli. Considering that most of the speech that we encounter everyday contains visual speech cues, it is imperative to move towards a greater understanding of how this information is used at both a behavioral and a neurophysiological level. Further, this research seeks to understand if populations with differential sensory experiences (e.g., congenitally deaf children who receive CIs) use audio and visual speech information to the same extent as those with typical development.

We presently build from predictive coding and neurocognitive accounts in order to create a comprehensive view of speech processing that incorporates naturalistic visual speech information. Based on the dual stream model of speech processing, we expect that visual speech information makes contact with auditory speech information in the ventral stream, in the sensorimotor cortex where input from multiple senses can be integrated (Baart & Samuel, 2015;

Hickok, 2022). The utility of this visual speech information may vary depending on factors such as the informativeness of the visual cue itself, the amount of noise in the environment affecting the fidelity of auditory information, as well as higher-level contextual cues such as preceding semantic information or speaker knowledge. To this extent, visual cues may or may not contribute to the predictions that are formed at phonological and lexical-semantic levels of processing. However it might be expected that in cochlear implant-users, visual cues play a greater role in both of these levels of processing given that auditory information is, at baseline, less reliable than it would be for a listener with typical hearing (TH). It may even be the case that in certain tasks or listening situations, cochlear implant users rely on strategic, predictive pre-activation processes in order to aid their comprehension of audiovisual spoken words (Kuperberg & Jaeger, 2016). This level of active engagement may not be necessary for TH listeners in the same situations, under the assumption that audiovisual speech provides more than enough reliable information in order to allow for passive, less effortful, bottom-up processing of spoken words.

Audiovisual Speech Processing in Typical Hearing Children

During the first year of life, infants show sensitivity to visual facial cues and can use this information to bootstrap their knowledge of language (Kuhl & Meltzoff, 1982; Teinonen et al., 2008; Yeung & Werker, 2013). Benefit provided by visual cues may even extend the developmental window in which infants are able to discriminate phonemes from non-native languages, prior to the onset of perceptual tuning for native language phonemes (Danielson et al., 2017). The salience of visual speech information seems to temporarily dissipate after the first year, as children begin to place greater reliance upon the auditory information provided by speech. Lalonde and Holt (2015) found that compared to four-year olds, three-year olds were less likely to use reliable visual information to improve their speech processing abilities and were less adept at AV speech integration. This finding and others suggests that after a period of sensitivity to facial cues in infancy, there is a period in early childhood in which visual speech information is not heavily relied upon in favor of auditory information. Children eventually begin to utilize information from both modalities in order to facilitate speech perception, though the specific age at which this emerges may vary across contexts and languages (Sekiyama & Burnham, 2008).

The benefits of congruent AV speech cues on processing are not as robust in children as they are in adults, and the emergence of these benefits has been linked to increases in experience in self-articulation of speech sounds, exposure to AV speech signals, and improved cognitive functions such as strategic use of lexical-knowledge (Desjardins et al., 2007; Eisenberg et al., 2000; Ross et al., 2011). There are mixed findings regarding at what time point in development and to what extent AV speech cues provide benefit to children's spoken language processing, likely due to a wide variety of tasks being utilized in these studies. Tasks involving simple matching may show an AV advantage in children as young as 3 or 4 years (Lalonde & Holt, 2015). Discrimination and recognition tasks, as well as speech processing in noise, may not show an AV benefit until later in childhood (Ross et al., 2011). This may be because these tasks place greater cognitive demands on children, such that they may have to form guesses about what words were heard if they have limited or unreliable sensory input within the task. Several studies have cited listening strategy as an important factor in speech perception capabilities (Eisenberg et al., 2000; Leibold & Buss, 2019; McCreery et al., 2017), suggesting that adult-like levels of speech perception capabilities, especially in noise, may not emerge until late childhood or older as children develop greater cognitive abilities such as learning to inhibit the perception of irrelevant acoustic cues (Nittrouer et al., 1993). Along with what is known about children's ability to integrate audio and visual speech cues, these patterns of findings provide support for the idea that AV speech integration follows a U-shaped developmental trajectory, such that the full use of all available speech input may be temporarily limited after infancy only to achieve adult-levels later on in development (Jerger et al., 2009; Lalonde & Holt, 2015).

Given the extensive behavioral research on AV speech processing and integration in both adults and children, there is growing interest to understand the neural mechanisms that support the development and efficiency of these processes. It is especially important to consider how AV speech perception may facilitate speech recognition in populations for whom the processing of auditory information may differ from that of typically developing children – specifically congenitally deaf children who make use of cochlear implants to comprehend spoken language.

Audiovisual Speech Processing in Cochlear Implant-Using Children

This section will specifically focus on studies that investigate multisensory speech processing in prelingually deafened CI-using children. For a comprehensive review of the literature on multisensory processing in adult CI users, please see Stevenson et al. (2017). Like children with typical hearing, children who rely upon CIs for auditory input can also show AV benefits to speech perception when compared to AO conditions (Lachs et al., 2001). Huyse and colleagues (2013) show this benefit at the level of phoneme processing. In this study, CI and age-matched TH children were presented with AV and visual-only speech stimuli, where the auditory input for the TH group was degraded to mimic the auditory input that the CI group received from their devices (an important and underutilized manipulation within this literature for assessing the influence of hearing through a CI). The results showed that both groups demonstrated similar levels of AV benefit at the phonological level, as evidenced through performance on a consonant identification task. Further, when visual speech input was degraded (preventing the use of lip-reading abilities) both groups could rely upon auditory speech information (Huyse et al., 2013). Recent evidence from Lalonde and McCreery (2020) suggests that compared to those with typical hearing, school age children with mild to severe sensorineural hearing loss were more adept at accessing both phonological-level and lexical-semantic level information from audiovisual speech. These children may benefit more from enhancements provided by visual speech cues than children with typical hearing, as evidenced by their performance on syllable detection and sentence recognition tasks.

AV speech processing skills may depend on children’s chronological age prior to implantation. In a study with TH and CI-using deaf infants and children, Bergeson and colleagues (2010) conducted an experiment with an intermodal preferential looking paradigm design, such that two silent videos of a speaker articulating the words “judge” and “back” were presented on different sides of a screen, while the auditory component of one of the words was played over speakers. A total of 16 trials (8 of each word) were presented in random order. TH (n = 20) and CI-using (n = 19) children ages 12 to 40 months participated in the study, and gaze directed at the congruent silent video was taken as evidence of ability to spontaneously integrate audio and visual speech information. The results found that while TH children appeared to spontaneously integrate audio and visual speech cues during the first half of trials in the paradigm (as evidenced by longer looking times to the matching video), CI children only demonstrated a preference for the matching video during the second half of trials. Additionally, performance on this task was better for children who were implanted later in life, possibly reflecting that these children were more attuned to visual speech information and were able to utilize these cues once they had some experience with the task (the second half of trials; Bergeson et al., 2010).

Other studies have reported similar findings that suggest visual aspects of speech may be especially salient for CI-using children who receive their implant later in childhood (Bergeson et al., 2001, 2005; Schorr et al., 2005). It should be noted that during the last twenty years, parents are consistently recommended to implant their children earlier and earlier. Whereas ten years ago a child may have been considered early-implanted at 24 months, now it is not uncommon for children to receive CIs as early as 9 months of age. Across many developmental CI studies, however, "late implanted" children can generally be considered children who receive their CI after approximately 36 months of age (Campbell & Sharma, 2016).

With this in mind, children who have had less reliable early auditory experiences than others (e.g., children with progressive hearing loss who do not receive CI intervention until after 36 months) may come to rely upon visual speech information as a means to improve their spoken language understanding. Schorr et al. (2005) tested CI-using children on AV speech that combined either congruent or incongruent speech cues. The incongruent AV speech stimuli were designed to elicit a McGurk effect (McGurk & MacDonald, 1976), which allows for detection of bimodal fusion. In the McGurk effect, the auditory speech information indicates the production of one consonant (e.g., "ka"), while the visual speech information suggests the production of another consonant (e.g., "pa"). The successful integration of this competing AV information often yields the perception of a third consonant (e.g., “ta”). Thus, children’s responses as to what consonant they heard during AV speech presentation is informative of their ability to integrate auditory and visual speech cues. On this task, early implanted CI-using children reliably perceived congruent AV speech similar to hearing controls. However, children who received their implant after age 2.5 years responded to incongruent AV speech such that they were biased towards the visual speech cues (Schorr et al., 2005). Similarly, Bergeson and colleagues (2001; 2005) found that CI-using children who received their implant after 53 months performed better on visual-only word and sentence comprehension tasks compared to children who received their implant before 53 months, further suggesting that late-implanted children may place greater reliance on or have enhanced processing capabilities of visual speech information.

However, in these studies by Bergeson et al. (2001; 2005), early-implanted CI-using children outperformed late-implanted CI-using children on these tasks in AO and AV conditions. Early-implanted CI-using children,who had earlier access to auditory information, were more adept at recognizing auditory and AV speech, which suggests that early auditory experience can influence CI-using children’s subsequent AV speech perception. Similarly, Stevenson et al. (2017) posit that early-implanted children obtain AV integration benefits similar to individuals with typical hearing, while those children implanted later are less likely to achieve the same degree of benefit. This suggests that there may be a sensitive period in development of brain networks that support AV integration.

Neural Correlates of Audiovisual Speech Recognition

The temporal nature of speech processing makes it a prime candidate for investigation using electrophysiological methods. Neural markers of audio-only and audiovisual speech processing can be studied using Electroencephalography (EEG) and Event-Related Potentials (ERPs). There may be differences in the time course or pattern of ERP waveforms elicited during the spoken word recognition process between CI-using and TH children. Described below are several ERP components that are related to audiovisual speech processing, beginning with the earliest-occurring (automatic, exogenous sensory responses) and leading up to the latest-occurring (components influenced by endogenous cognitive processes). We review the components as they have been described in the adult literature, as well as developmental accounts of these components.

P1. The visual P1 component, which is a positive-going peak that occurs about 100 ms after stimulus presentation, is typically thought of as a neural marker of visual processing; thus, this component is particularly important for examining neural responses to visual speech cues. More specifically, the P1 reflects largely exogenous, automatic influences on the visual system (Kaganovich et al., 2016). It is sensitive to the sensory properties of visual objects (e.g., stimulus contrast), and is maximal over occipital sites. The amplitude of the P1 is shown to be modulated by visual attention (Luck et al., 2000; Mangun & Hillyard, 1990), such that voluntarily directing attention to a location within a visual field can increase the P1 amplitude to visual targets that appear within this location relative to targets in unattended locations (McDonald & Green, 2008). Studies of post-lingual deaf adults who received CIs in adulthood have reported decreased latencies of the P1 component compared to hearing controls (Hauthal et al., 2014; Sandmann et al., 2012) which is interpreted to reflect more efficient processing of visual stimuli.

The P1 in Development. In the context of congenital deafness, reports of modulation of the visual P1 have been mixed. For example, in pediatric populations responses to simple visual stimuli such as static checkerboards have not revealed prominent latency or amplitude differences between typically hearing and CI-using children (Corina et al., 2017; Liu et al., 2017), though visual onset responses to flashing checkerboards do evoke larger visual P1 amplitudes in CI-using children compared to controls (Corina et al., 2024).

There are not many comparisons of visual P1 responses between typically hearing and CI-using children to ecologically valid stimuli, such as AV speech stimuli. Pierotti and colleagues *(in preparation*) compared P1 responses from TH and CI-using school-aged children as they were presented with short videos of a speaker saying a word, as part of a larger word-picture priming paradigm. The results of this study found that compared to TH controls, CI-using children had larger P1 responses but that there were no latency differences between groups. Further, P1 responses to static images of the speaker's face did not yield any group differences. Together, these findings suggest that visual cues presented in conjunction with speech in the form of AV videos may be particularly salient for CI-using children, above and beyond static visual faces alone (see Corina et al., 2024, for a greater discussion of the engagement of the saliency network in CI-using children during the processing of visual stimuli).

N1. The N1 is a widely-distributed, frontal negative component that occurs roughly 100 ms after the onset of auditory stimuli and is considered, along with the P2, to reflect early automatic sensory processing (Näätänen & Picton, 1987). The amplitude of the N1 is modulated by acoustic features such as frequency or intensity, suggesting that this component reflects selective attention to basic auditory features and detecting patterns in auditory signals. These steps in auditory processing are foundational for subsequent steps that occur in the process of recognizing speech. In a picture-spoken word priming task, it was shown that adults had more negative N1 amplitudes in response to mispronounced words and pseudowords, supporting the idea that the N1 is affected by acoustic differences in speech (Duta et al., 2012). Providing further evidence that the N1 is sensitive to differences in speech sounds, Coch and colleagues (2002) found that latency of what they called the N120 was shorter for pairs of words that did not rhyme compared to rhyme pairs. Even further, the N1 has been shown to be sensitive to semantic context, where the amplitude of the N1 was reduced in response to semantically-associated spoken word targets in cross-modal priming study (Getz & Toscano, 2019). In the same study, Getz and Toscano show that semantic context can further influence the encoding of speech sounds during the N1 window, where targets that had ambiguous voice-onset times were processed more similarly to either voiced (larger N1 amplitude) or voiceless (smaller N1 amplitude) targets depending on the semantic context established by the prime. These findings provide evidence for the influence of top-down contextual information on early auditory processing during the spoken word recognition process.

The N1 in Development. The N1 is not often visible in young children, especially if they are passively listening to repetitive speech stimuli[1]; more active engagement with less repetition and longer interstimulus intervals (ISIs) can draw out a N1 in developmental populations (Bonte & Blomert, 2004). Given the developmental trajectory of the N1 in which this component becomes increasingly prominent from early to late childhood, its presence in response to auditory

[1] The mismatch negativity (MMN) is often used rather than the N1 to measure detection of acoustic changes without attention in very young children and infants (see Cheour et al., 2000). We do not discuss the MMN in the present research because it is more relevant to the present hypotheses to understand the emergence of the Auditory N1 (as a marker of a mature auditory processing system) in our age group and special populations.

information has also been thought of as a neural index of a mature auditory processing system (Henderson et al., 2011). Developmental populations in which early auditory processing mechanisms may be compromised, such as children with dyslexia, Developmental Language Disorder (previously referred to as specific language impairment or SLI), or those who use cochlear implants, have shown aberrant N1 responses (e.g., longer latencies or less pronounced waveform patterns) when compared with age-matched typically-developing peers (Bonte & Blomert, 2004; Corina et al., 2022; Malins et al., 2013). Crucially, however, group differences in the N1 do not seem to be directly related to downstream comprehension differences of spoken words, suggesting that even with the presence of a deviant N1 response, speech recognition can be successful.

In children between ages 6-12 years with typical language skills, the latency and amplitude of the N1 have been shown to decrease in the presence of AV as opposed to auditory-only or visual-only speech, but this effect appears to be sensitive to attentional load and to the congruity of the audio and visual cues (Alsius et al., 2014; Knowland et al., 2014; though also see Brunellière et al., 2020 for a contradictory N1 result in adults who processed AV sentences). This attenuation effect has been attributed to using visual speech cues in order to predict the onset of auditory signal, such that this prediction can easily occur when cues from both modalities are consistent with each other (Kaganovich et al., 2015, 2016).

N280. If initial auditory processing occurs in the N1 time window, then expectations about upcoming speech are thought to be processed in the next window, associated with an N280 component. Unexpected speech sounds have been shown to elicit a pronounced peak in this window, sometimes referred to as the phonological mapping negativity (PMN).[2] This negative-going deflection appears 250-300 ms after spoken word stimuli, often maximal over fronto-central sites. Because it is only demonstrated in auditory word tasks and not written word tasks, it is proposed that this component reflects top-down contextual influences on a bottom-up phoneme mapping process. The N280 is thought to index prelexical mechanisms of spoken word recognition because it can occur for phonological mismatches that occur in either words, pseudowords, or nonwords, suggesting the negativity of the N280 arises before lexical status is assigned (Duta et al., 2012; Newman et al., 2003). The "N280 effect" appears to be exaggerated in the presence of AV speech when compared to AO speech, such that expected incoming words show a reduction in N280 amplitude whereas incongruent or unexpected words show an increase in N280 amplitude (Brunellière et al., 2020). This is possibly due to the complex interactions of phonological information coming from both the auditory and visual domains further restricting a set of phonological competitors. For example, when recognizing the word "nose", phonological competitors "rose" and "hose" can be ruled out even more quickly based on the combined speech cues from both phonemes and visemes than based on information from one modality alone, thereby reducing processing costs when incoming words are consistent with predictions but increasing costs when incoming words are incongruent with predictions.

With the goal of understanding the influence of phonological similarity on the time course of ERPs associated with spoken word recognition, Desroches and colleagues (2009) presented adult subjects with pairs of pictures and spoken words in a matching task. Trials

[2] There are a few names for the negative-going component and its corresponding modulation that occurs around 280 ms in response to spoken language stimuli, but for the present study we will be referring to this component as the N280, and its modulation based on phonological congruity will be referred to as an N280 effect.

consisted of the match condition, but also mismatch conditions of rhyme, word initial cohort, and unrelated spoken words. Enhanced negativity in the N280 window was observed in unrelated and rhyme trials, but not in the match trials. The word initial cohort condition did not elicit an N280 because the phonological onset was consistent with what was expected (e.g., the phonological violation occurred at the end of the word). The subsequent significant negative deflection observed in a late N400 window for word initial cohort trials is proposed to reflect an additive negative effect of violated phonological and semantic expectations that are delayed by the later onset of the mismatch information. The findings of this study suggest that prelexical phonological processing and lexical processing of word meaning are separable but interactive in their contributions to the recognition of spoken words.

The N280 in Development. As for developmental studies investigating the N280, Bonte and Blomert (2004) administered a spoken word lexical decision task to school age children with typical development or dyslexia. The goal of their study was to understand implicit phonological processing differences between groups during spoken word recognition, using ERPs associated with prelexical processing (specifically the N1 and a component similar to the N280 they labeled the N2). In this study, children performed an auditory lexical decision task in which pairs of words and nonwords were alliterations (shared the first two phonemes) or did not have any phonological overlap. Dyslexic children showed greater N2 negativity compared to controls, which under an N280 account, may suggest more effortful phonological processing for this population compared to TH children. Furthermore, replications of Desroches et al. (2009) in school-aged children with dyslexia (Desroches et al., 2013) and Developmental Language Disorder (Malins et al., 2013) demonstrate that the N280 effect occurs under the same context and with the same modulations (e.g., more negative amplitude for phonological mismatches) for these developmental populations, as well, though there may be slight deviations in each population's sensitivity to phonological violations (e.g., rhymes).

P300. Often implicated with endogenous processing of improbable information, the P300 (or P3b) is a positive going peak that is maximal over centro-parietal sites between 250-500 ms after a stimulus. While the P300 is not typically considered a language ERP component, it is discussed here due to its relevance in processing unlikely information. Typically demonstrated in auditory oddball paradigms, less probable information will evoke more positive P300 amplitudes; this relationship has been proposed to reflect information processing costs, which may vary according to individuals' working memory capacity, subjective probabilities of the stimuli, and preceding stimulus information that could influence prediction (Levi-Aharoni et al., 2020; Van Petten & Luka, 2012). This interpretation of the P300 has been discussed in research with both adults and children (Brydges et al., 2014; Polich, 2007).

Notably, within the literature on working memory updating and the processing of improbable information, the P300 has been studied with respect to speech processing in both adult and child recipients of CIs. Beynon and Snik (2004) summarized the results of several studies in which adult CI users showed prominent P300 responses to various speech stimuli, though with slower latencies and reduced amplitudes than hearing controls, possibly due to adult CI users' varied sensory experiences. Children with CIs, however, only elicited reliable P300 responses to pure tone tasks, and not to speech. This was taken to suggest that one of the hypothesized subcortical generators of the auditory P300, associated with phonetic processing, may be immature in these children (Beynon & Snik, 2004). More recently, Abarahamse and colleagues (2021) observed that in young adult CI users, duration of deafness and performance on a speech perception task were correlated with P300 amplitude in an auditory oddball paradigm. In other words, shorter periods of auditory deprivation and better consonant discrimination (/ba/ vs. /da/) were associated with larger P300 amplitudes (Abrahamse et al., 2021).

The studies mentioned here are only a small sample of the work related to an auditory P300/P3b component. In the context of the present study, we aim to emphasize the P300 as an information-processing component. Given the evidence that the P300 reflects an updating of working memory and disconfirmation of expectations, we may infer in the present study that the P300 component corresponds to a step in spoken word recognition that registers unexpected information (Van Petten & Luka, 2012). With respect to other speech processing-related components, the P300 response may mask or overlap with the occurrence of N280 or N400 responses. A prominent P300 response, especially if present in place of an N280 component, may indicate that a listener is forming expectations based on the broader context of the task (e.g., likelihood of a target being a match or mismatch). On the other hand, if a listener forms predictions on a trial-by-trial basis, predictions may be informed by incoming sensory input (e.g., visual and auditory cues), along with more immediate context such as a preceding prime stimulus. Under these conditions, we may expect the listener is attending to phonological information of primes and targets, and could be more likely to display an N280 response.

N400. The N400 is a well-studied negative-going component that is maximal from about 300-500 ms after presentation of meaningful stimuli (including spoken words and pictures; Kutas & Hillyard, 1980; Swaab et al., 2012). This component is maximal over central and parietal sites, though can sometimes appear over frontal sites especially in response to pictorial stimuli (Barrett

& Rugg, 1990). The modulation of the N400 component observed in congruent and incongruent contextual environments is referred to as the N400 effect. This component is commonly associated with registration of the semantic content of stimuli, but the specific process that contributes to this semantic registration is subject to debate (Kutas & Federmeier, 2011; Lau et al., 2008; Nieuwland, 2019).

One possibility is that the N400 wave is an index of ease of lexical access, such that the amplitude of this component is more negative in instances where lexical items are harder to retrieve (e.g., low frequency words). A lexical access view holds that the N400 effect reflects facilitated activation of features of the long-term memory representation that is associated with a lexical item (Lau et al., 2008) or prediction error at the lexical-semantic level (Eddine et al., 2024).

Another possibility that is more domain general than the lexical access view - in that it can account for pictorial and other non-lexical N400 components - is that the N400 reflects how easily semantic content can be integrated with prior context (Swaab et al., 2012) . Under an integration account, the N400 effect reflects a combinatorial process in which the ease or difficulty of integrating a current stimulus to a prior context modulates the magnitude of the effect. A concrete example of this can be found in priming tasks, where a prime (e.g., a picture) establishes a context in which a target (e.g., word) can either be semantically consistent or inconsistent with the meaning of the prime. Semantically consistent or congruent targets are presumably easier to integrate with the established context than an incongruent target, and would therefore evoke a reduced N400 amplitude.

As discussed throughout this introduction, there is a growing body of research that argues that language comprehension is not a passive process, but instead we are actively constructing meaning and generating expectations about how language will unfold (Brothers et al., 2015; Federmeier & Kutas, 1999; Kuperberg & Jaeger, 2016; Maess et al., 2016). Thus, the amplitude of the N400 may reflect implicit prediction error, such that the amplitude of this component will be greater (more negative) when there is more discrepancy between the meaning of the predicted word and the meaning of the actual target (Rabovsky & McRae, 2014). Much of this research has been inspired by or connected to Bayesian principles of prediction, in which listeners incorporate probabilities based on previous experiences into their anticipations about upcoming linguistic information (see earlier discussion of Shortlist B and predictive theories of spoken word recognition; Delaney-Busch et al., 2019). Delaney-Busch and colleagues (2019) used a Bayesian learning model to account for trial-by-trial variance in the N400 amplitude under semantic priming conditions, whereby the model would update the probability of encountering a semantically congruent or incongruent target after every trial. This follows the logic that semantic priming effects are larger in blocks of trials that have a greater proportion of congruent versus incongruent prime-target pairs. Models such as this provide compelling evidence that listeners can account for the statistical structure of their environments and that this can inform predictions about upcoming words. Listeners can also attenuate the strength of their predictions when their predictions are less reliable or more likely to fail (see discussion of *rational adaptation* in Ness et al., 2021). It is less clear, however, if there are individual differences in how costly it is to use these active predictive strategies during spoken word recognition, or how prediction may emerge in special developmental populations such as deaf CI-using children.

While the previous paragraphs have outlined the influence of lexical information, broader context, and active predictions on the N400, it is of relevance to note that amplitude and latency of the N400 can be susceptible to phonological information, such as word-initial phonological overlap and rhymes. Connolly and Phillips (1994) noted a late shift in the latency of the N400 during a sentence listening task when the target word had a congruent phonological onset (e.g., "the gambler had a streak of bad *luggage*" instead of "*luck*"). A delayed N400 with increased negativity (relative to match, unrelated, and rhyme conditions) was also observed by Desroches et al. (2009) when subjects were presented with word-initial cohort pairs of pictures and spoken word stimuli.

Rhymes, on the other hand, have been shown to reduce the magnitude of the N400 effect (Desroches et al., 2009; Praamstra & Stegeman, 1993). Consistent with lexical accounts of the N400, there appear to be advantages to lexical access when words are primed with phonologically-similar words such as rhymes, as indexed by an attenuated N400 amplitude. Praamstra and Stegeman (1993) measured the N400 during a spoken word rhyme decision task, using words and non-words. Their results showed that the N400 waveform was more attenuated when pairs of words rhymed than when they did not rhyme, but that this pattern of results did not occur for the non-lexical stimuli. N400 reduction to rhymes has been attributed to the idea that pre-activation occurs for rhyme lexical candidates, which in turn facilitates recognition of these candidates at a lexical stage of processing that occurs in the N400 window. Additionally, Desroches and colleagues (2009) found that N400 negativity to unrelated picture-word pairs was more sustained than negativity to rhyme picture-word pairs, based on analysis of two separate N400 time windows (e.g., an N400 window from 310-410 ms and a late N400 window from 410-600 ms). Others have reported a late-occurring anterior negativity in response to rhyming word stimuli, which has also been reported to be lateralized to the left hemisphere more so than an N400 response to non-rhyme spoken words (Coch, Grossi, et al., 2002). Coch and colleagues (2002) reported this pattern of rhyme N400 responses in children as young as age 7, suggesting that the neural processes underlying this component may be mature by middle childhood. More recent evidence from Andersson and colleagues (2018) shows that when presented with pairs of spoken non-words that either rhymed or did not rhyme, preschool-aged children also elicited a lateralized anterior negativity to rhyming stimuli. In contrast, non-rhyming stimuli elicited a more negative posterior N400 response, consistent with the adult literature.

As suggested by Praamstra and Stegeman (1993), it can be difficult to dissociate between the N280 and the N400; the N280 has even been referred to as an "early N400" or "phonological N400". Van Petten et al. (1999) argue that semantic integration can begin even if there is only partial information about the word form, which may explain why the negativity occurs in the early window that has been associated with phonological processing. According to this argument, this negativity is not a reflection of phonological mapping or the error detection of a phonological mismatch but is instead initial semantic integration processes that are associated with the N400. Connolly and Phillips (1994), however, provide evidence that the N280 can be separated from semantic processing. Subjects participated in a sentence listening task where the final word of the sentence was either a) the highest Cloze probability word, b) a word that shared the initial phoneme of the highest Cloze probability word but was semantically incongruous with the sentence context, c) a word that differed in initial phoneme from the highest Cloze probability but was semantically congruous, or d) a word that was both semantically and phonologically anomalous. A more negative N280 was present in trials where there were initial phonological violations of the expected word form (e.g., condition c), regardless of if there was also a semantic violation (Connolly & Phillips, 1994). Further, the latency of the N400 was shifted in condition b where the initial phoneme of the target word overlapped with the highest Cloze probability word, suggesting that the point of disambiguation between semantically expected and unexpected words may shift the time course of neural waveforms that index difficulty of semantic processing.

The N400 in Children with Cochlear Implants. While our understanding of the precursors and mechanism of the N400 effect are evolving, it should be further noted that current hypotheses of causal mechanisms underlying N400 effects are largely founded on findings from adult language processing and the degree to which these alternative hypotheses are applicable to pediatric and special populations remain understudied. Indeed, while the N400 effect has been consistently elicited in typically developing verbal children in response to spoken words and pictures (Coch, Maron, et al., 2002; Cummings et al., 2009; Friedrich & Friederici, 2004; Henderson et al., 2011), there are only a few studies that have investigated the N400 effect in CI-using children.

Kallioinen and colleagues (2016) used a spoken word and picture semantic priming task to elicit N400 responses from 30 deaf and hard of hearing children aged 5-7 years, 15 of which had at least one CI. The authors predicted that larger N400 mismatch effects would occur for hearing children compared to deaf children, under the assumption that semantic discrimination would be easier for groups with better hearing. Unexpectedly, the children with CI demonstrated larger effects in response to between-category semantically mismatched items than hearing controls and children with hearing aids. Behavioral results did not suggest that the CI-using children had better semantic discrimination; additionally, the timing and magnitude differences observed between groups led the authors to tentatively conclude that children with CI may have less precision in semantic processing, or a stronger reliance on predictive processing; this conclusion, however, was not explicitly related to the N400 (Kallioinen et al., 2016).

In contrast to the previous study's findings, Bell et al. (2019) did not observe group differences in N400 effects between CI-using children and typical hearing controls on their spoken word-picture priming paradigm, despite their predictions that semantic integration would be more difficult for the CI-using children. Behavioral measures, however, indicate that CI-using children performed significantly worse than hearing children on tasks related to spoken language comprehension. The authors' interpretation of these results is that though spoken language difficulties occur for children with CI, these difficulties are not represented at a neural level by the N400.

Most recently, Pierotti et al. (2021) tested 29 CI-using children and 19 TH controls ages 2-10 years on a passive audiovisual word-picture priming task. Videos of speakers saying words aloud were followed by either semantically congruous or incongruous picture targets, with an interstimulus interval of 400 ms. Children were instructed to listen/watch passively as continuous EEG was recorded. They found that while the groups showed no difference in N400 amplitude to semantically congruent pictures, the CI-using children's N400 mismatch effects to semantically incongruent pictures were more negative than TH controls. Interestingly, the mismatch effects began in an earlier window than expected (around 200 ms); at the time, this was taken to suggest that there may be attentional or strategic differences in semantic integration amongst groups. More specifically, the CI group may have been evoking an active prediction-based strategy (consistent with the discussion of Kallioinen and colleagues' findings), in which incongruent semantic information contained in the picture targets may have elicited a greater implicit prediction error for this group than the control group. Relative to controls, CI children may have been more committed to their predictions given that the primes were audiovisual in nature, offering multiple speech cue modalities from which predictions can be built. Recall that there is research showing that visual cues may be particularly beneficial to CI-using children during speech processing (Schorr et al. 2005). The control group on the other hand, may have been less committed to their predictions or perhaps were using more automatic or passive processes during integration of the semantic content of the word primes and picture targets.

Further consideration of the pattern of results reported by Pierotti and colleagues has led to the question of how the N280 may have been reflected in the CI-using children during this paradigm. It is unlikely that the early negativity demonstrated by the CI group around 200 ms is reflective of phonological processing, because the ERPs were time-locked to the onset of the picture targets which do not contain any phonological information. Despite this, it is important to consider how the combination of audio and visual cues provided by the word primes may have influenced subsequent processing of the picture targets, especially given the timing of presentation of prime and target stimuli. A 400 ms delay between the offset of the word and the onset of the picture may have led to lingering effects in the picture window, especially in some of the youngest subjects. While the present work has outlined much of the research that shows attenuation of speech-related ERP components in the presence of AV as opposed to AO speech, it has also been suggested that for the N280 specifically, AV speech may enhance the elimination of phonological competitors given information from multiple modalities (Brunellière et al., 2020). It is not clear if the early negativity shown to pictures is indeed a lingering N280, especially because there was no evidence of a lingering late N400 in the results of Pierotti et al., 2021. However, it remains that this component has yet to be fully explored in this population or in response to AV speech. Thus, more work is needed to understand what interactions phonological and semantic information may play during the neural processes supporting AV spoken word recognition in deaf CI-using children.

Present study

The present research compares neural patterns of spoken word recognition through the use of a picture-word paradigm between children with typical hearing and those who hear through a CI. If group differences exist, we hope to reveal them by manipulating phonological and semantic congruity of picture and AV spoken word stimuli, as has been done in previous adult and developmental studies investigating spoken word recognition (Desroches et al., 2009, 2013; Malins et al., 2013). The current research utilizes the same conditions of picture and word congruence and incongruence as Desroches and colleague's (2009) original paradigm (*matches*, see cake and hear "cake"; *unrelated mismatches*, see cake and hear "seal"; *rhyme mismatches*, see cake and hear "rake"; and *word initial cohort mismatches*, see cake and hear "cage"). However, we extend the previous work by using spoken word targets that include both audio and visual speech cues through the use of video stimuli. We compare AV word recognition between age-matched CI-using and TH children, with a focus on the following ERP components: the visual P1, the auditory N1, the N280/P300, and the N400.

We entertain several hypotheses: first we expect the CI group to show faster onset and/or more positive occipital P1 amplitudes than the TH control group in response to AV spoken word stimuli, as demonstrated previously (Corina et al., 2024; Pierotti et al., *in preparation*). This hypothesis is based on the evidence suggesting that visual speech cues may be particularly salient for CI-using children who find this modality more reliable than the auditory modality. Next, following the idea that CI-using children may have less mature auditory processing systems compared to TH controls, we hypothesize that the CI group will have longer N1 latencies and/or reduced N1 amplitudes relative to the TH control group, regardless of condition. Similar patterns of effects on early auditory-evoked components have been reported in prior studies of CI-using children (Corina et al., 2017, 2022).

Further, with respect to the N280, we expect to replicate previous findings using this study design such that we may observe phonological mismatch condition differences: rhymes and unrelated mismatches are predicted to have more negative N280 amplitudes than matches and word initial cohort mismatches due to the initial phonological onset incongruity that occurs between rhymes and unrelated pairs of stimuli. As for group differences in the N280, our predictions are less clear given that this component has not been studied in children with CI. While Bonte and Blomert (2004) observed more negative amplitudes in the N280 timeframe for dyslexic children when compared to controls, other studies have not observed differences between typically developing children and groups of children with language-related impairments (Desroches et al., 2013; Malins et al., 2013). To the extent that phonological processing differs in CI-using children, we may observe more negative N280 amplitudes as a reflection of more effortful mapping of phonemes, particularly on unrelated and rhyme mismatch conditions in which the target word phonemes are incongruent from the phonemic expectations established by the picture prime.

We predict that there will be condition differences in N400 amplitude and latency similar to what has been demonstrated by Desroches et al. (2009) and subsequent replications. Semantically unrelated mismatched word targets are anticipated to evoke a more negative N400 amplitude than semantically matched word targets. Further, N400 amplitudes in response to unrelated mismatches are predicted to be more negative than matches and rhyme mismatches.

Rhyme N400 responses are often similar in magnitude to match N400 responses due to the facilitation in lexical processing of rhymes as a result of spreading lexical activation. The latency of the N400 is expected to be later for word initial cohort mismatches given that there is a later point of disambiguation between match targets and word initial cohort targets; therefore we intend to measure a "Late N400" at a window beyond what is classically observed for the N400 in order to capture this shift. Once this point of mismatching is reached, we predict that the word initial cohort mismatch N400 amplitudes will be even more negative than those of the unrelated mismatches, under the assumption that the prediction error at this later stage of processing will lead to even greater processing difficulties than for words that do not share phonological onset with the prime. It is less apparent that this anticipated pattern of responses for each incongruent condition will occur in the CI group, given that we do not yet understand the influences of phonological and semantic congruence on spoken word recognition in this population. CI-using children, similar to children with Developmental Language Disorder or dyslexia reported by previous studies (Desroches et al., 2013; Malins et al., 2013), may have difficulty differentiating rhymes from matches, thus rhyme mismatch and match N400 responses may be indistinguishable within this group. Additionally, CI-using children could be slower to resolve word-initial cohort mismatches, which could further extend the latency of this condition's N400 response compared to TH controls. Pierotti and colleagues (2021) have also previously reported significantly more negative N400 responses to unrelated mismatches in CI-using children when compared to age-matched TH controls. It is possible that this difference between groups (previously attributed to the use of a more active semantic processing strategy and greater reliance upon predictive processing in CI-using children) could appear in the present study.

Finally, we predict that there will be associations between neural measures and other subject characteristics such as demographics and behavioral responses. To the extent that CI-using children who have greater experience in sound[3] display more similar neural responses to children with typical hearing, we may expect that a greater Time in Sound (TIS) is associated with less positive P1 peak amplitudes and more negative N1 peak amplitudes in response to AV speech. Both of these associations would suggest that CI children are showing greater early sensory responsivity to the audio information of the AV speech as opposed to the visual cues. We may also see developmentally-driven changes in early sensory component peak amplitude and latency: as visual and auditory systems mature and become more efficient, the neural markers of these processes may reflect this in smaller peak amplitudes and shorter latencies. Another factor that we consider is task accuracy and response time. As we have introduced, children's strategy use during the task may influence the ERP components that emerge, especially with later cognitive components. We may find that children with better accuracy, regardless of which group they are in, may demonstrate more positive componentry similar to a P300 compared to negative componentry indicative of an N280. Further, if we take the presence of a P300 to reflect the use of an efficient, global strategy in which expectations are adjusted based on task demands, we may find that the children who are the most accurate on this task are the ones who are using this strategy. The last correlation we will examine is a potential association between the amplitude of the N400 with receptive and expressive vocabulary. Across both groups, it may be true that children with larger vocabulary sizes are more attuned to semantic properties of words

[3] "Experience in sound" can be defined by the age at which a child receives their cochlear implant (Age of Implantation; AOI) and by duration of CI use (Time in Sound; TIS). Though there are many other factors that can contribute to CI device "success", it is typically expected that earlier ages of implantation and greater duration of CI-use lead to better hearing and language outcomes for this population (Sharma et al., 2002).

(Schneider et al., 2023), and may demonstrate smaller N400 effect sizes to semantically incongruous word targets as a reflection of less effortful semantic retrieval.

Methods

Participants

A group of 13 children with cochlear implants participated in the study after their caregivers provided written consent. This sample consisted of 8 boys and 5 girls, with a mean age of 127.69 months (*SD* = 19.64; range 98–163). Mean age of first implantation was 28.62 months (*SD* = 27.26; range 7–82). The average Time-in-Sound (TIS) for the children in this sample was 99.08 months (*SD* = 34.26; range 16–133). Attention Deficit Disorder (ADD) was reported by parents for three of these children, and one child was born with normal hearing and diagnosed with progressive hearing loss at 18 months. Three of these children had prior or concurrent exposure to ASL at the time of their participation in this study. **Table 1** presents characteristics and demographics of the CI-using subjects in this study. This study was conducted in accordance with the recommendations of the University of California, Davis, Institutional Review Board (protocol 235840).

Table 1. Cochlear implant-using subject demographics, n =13.

Subject	Age*	Gender	Age at First Implant*	Time in Sound*	CIs	Interval Between CIs*
1	99	F	81	18	2	0
2	103	F	13	90	2	0
3	104	M	7	97	2	0
4	114	F	12	102	2	0
5	121	M	12	109	2	0
6	123	M	9	114	2	3
7	133	M	12	121	2	0
8	133	F	18	115	2	17
9	134	F	31	103	2	10
10	144	M	15	129	2	5
11	144	M	79	65	1	*N/A*
12	147	M	14	133	2	0
13	164	M	62	102	1	*N/A*

* Age/time variables are presented in months.

Comparison Group. For comparison with our CI-using group, a group of age-matched children with typical hearing, as reported by their parents, were recruited for the study. This group consisted of 22 children (12 boys and 10 girls) with an average chronological age of 123.64 months (*SD* = 19.76; range 89–159).

Stimuli

Commonly-used monosyllabic noun concepts originally used by Desroches et al. (2009) were repurposed for the stimuli in the present study[4]. Concepts were represented in both picture prime and audiovisual word target form. Picture primes consisted of color stock photographs of objects, which were verified in pilot testing by eight children of the desired age range for recognizability and naming consistency. Only pictures that achieved above 90% naming consensus were used.

[4] Some words were updated to be more age-appropriate or recognizable for children.

Audiovisual word targets were created through video recordings of a female native English speaker saying the word label for each concept. Videos were recorded using a SONY HXR-NX5U camera and microphone against a green screen, which was replaced with a uniform blue background (Final Cut Pro Version Pro 10.6.3, 2022). Example prime and target stimuli for each condition are shown in **Figure 1**. The sound from each video was stripped and saved as a separate audio file so that we could separately account for the onset of both the auditory and visual information from the targets (in other words, we were able to time-lock our ERPs to the onset of either the video or meaningful audio in each trial). Auditory files were resampled to 48,000 Hz and normalized to a loudness of -23.00 LUFS (Adobe Premier Pro Version 22.5.0, 2022). The sound onset across all audio files was kept as consistent across stimuli as possible, ranging from 180 to 284 ms after the start of the file (*M* = 238 ms, *SD* = 17.9). Some variability in sound onset remains, however, as there is natural variability in the timing of onset of mouth/visual speech information and auditory speech information (refer to /m/ and /h/ sound example provided in the introduction). We determined it was more important to ensure the speaker's mouth was closed at the beginning of the video than to ensure a completely even sound onset time across stimuli. The duration of the video primes ranged from 1201 to 1902 ms in duration (*M* = 1537 ms, *SD* = 134.4 ms). Within these videos the audio portion of the spoken word lasted on average 782 ms (SD = 146.9 ms).

Figure 1. Example stimuli

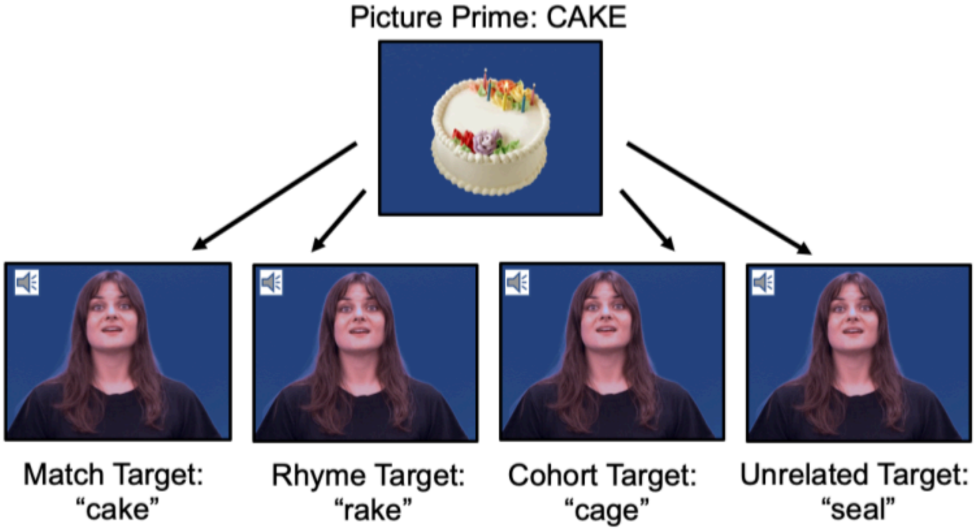


Figure 1. Example of a picture prime and subsequent audiovisual word targets of the four experimental conditions.

We also collected measures of word frequency and identified the initial viseme of each word. For each initial viseme, we calculated a measure of “viseme distinctiveness”, in which the likelihood of a viseme corresponding to a given phoneme is a ratio of how many other phonemes share the same viseme (e.g., /b/ and /p/ share a viseme, and therefore this viseme has a distinctiveness of 1 viseme divided by 2 phonemes or 0.5). All of the characteristics of the words used and their corresponding AV stimuli features are listed in **Appendix A**.

Additional visual stimuli in the form of static speaker images were created to be included in the paradigm. These static speaker images consisted of screen shots from the target videos in which the speaker was mid-production of the target word. Static speaker images have been used in previous studies to serve as a baseline for visual processing of speech-salient facial stimuli (see Pierotti et al., *in preparation*, for comparison of neural responses to static speakers and audiovisual speech).[5]

Procedure

The experimental procedure began with a naming consistency task in which children were shown all of the images to be used in the main part of the experiment. Children were prompted to name each picture using one word. If a child used a different name than the label that was used in the experiment (e.g., “sofa” instead of “couch”), experimenters first prompted the child to use a different label if they knew of one. If the child did not know of another label, the experimenter would name the picture using the experiments’ label and would mark whether or not the child recognized this word. This was done to ensure naming consistency among children on pictures that may have competing lexical entries, and also to identify trials in which the child would not know the correct answer given that they did not know the word, so that these trials could be analyzed separately.

The main experiment consisted of 186 trials, half of which were match trials. In match trials, the represented concept in the picture prime and audiovisual word target were consistent (e.g., picture of a cake followed by the spoken word “cake”). The other half of the trials were of either a rhyme condition (e.g., picture of cake followed by the word “rake”; 31 trials), a word-initial cohort condition (e.g., picture of cake followed by the word “cage”; 31 trials), or an unrelated condition (e.g., picture of cake followed by the word “seal”; 31 trials). Each concept was presented in both picture and audiovisual word form twice: once in a match condition and once in one of the other incongruent conditions. We counterbalanced the presentation of trials

[5] Briefly, the results of this study showed that CI users and TH controls did not differ in their response to static speaker images, but CI users showed enhanced P1 responses indicative of greater visual reactivity to AV speech stimuli relative to TH controls. This study is the first to differentiate between neural responses associated with speaker face processing and multisensory speaker input in the CI-using developmental population.

across two versions, such that in one version a concept was presented in a match trial first and in the other version the concept was presented in an incongruent trial first.

Refer to **Figure 2** for a detailed representation of a trial schematic. Each trial consisted of the presentation of a fixation cross for 250 ms, followed by the picture prime presented for 2000 ms and the audiovisual word target. The targets consisted of the video files and audio files being presented simultaneously to create audiovisual words. We used an interstimulus interval of 200 ms and an intertrial interval of 800 ms. After the target presentation, a mouse prompt appeared that instructed them to click left or right to indicate if the picture and word matched. Left or right clicks for indication of match trials were counterbalanced across versions of the experiment.

Figure 2. Trial schematic

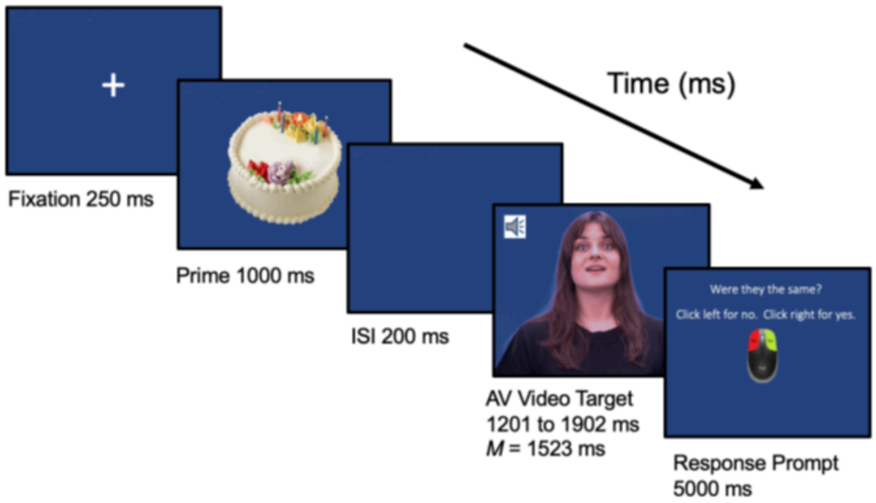


Figure 2. Timeline schematic of a trial during the experiment. The experiment consisted of 186 trials, 93 of which were match, 31 rhyme, 31 cohort, and 31 unrelated. Six self-paced breaks were offered to subjects evenly throughout the experiment.

Prior to the beginning of the experiment, children were given the instructions that they were going to be shown pictures and videos of a woman naming the pictures, but that sometimes the woman would name the picture incorrectly. They should listen carefully to determine if she had used the matching name for the picture, and to respond as quickly and accurately as possible when they knew if the picture and the word matched each other. They were allowed to respond before the video offset instead of waiting for the mouse prompt. Unless a response was made to terminate the trial, the mouse prompt lasted for 5000 ms before disappearing. Nine practice trials of either match or unrelated conditions were used at the beginning of the experiment prior to recording so that participants could familiarize themselves with the procedure. These practice trials were used in addition to the 186 main experiment trials and were not included in the present analysis.

All visual stimuli were presented on an LCD monitor with a uniform blue background, and auditory stimuli were presented using two stereo speakers (AudioEngine A5+) with a sound level set to 65dB(A). Evenly interspersed through the main experiment were six self-paced breaks in which children saw on-screen instructions to rest their eyes and click the mouse when they were ready to continue. Immediately after the response to end the break, a static speaker image was presented for 1000 ms. These six static speaker image events were included to serve as a baseline for visual responsivity to facial information without the presence of speech. No response was required to static speaker images.

In addition to collecting EEG data in response to the picture-word matching paradigm, all children completed the Expressive and Receptive One Word Picture Vocabulary Tests to gauge their expressive and receptive vocabulary (EOWPVT and ROWPVT, respectively). Almost all children (n = 34) completed these tests immediately following the EEG task, and the other child completed the vocabulary within 6 weeks of EEG testing. The entire session (EEG and behavioral testing) lasted about one hour.

EEG Analysis

Continuous EEG data was collected from 22 electrode sites, using the standard 10/20 system with the Biosemi Active Two System (Biosemi B. V., Amsterdam, Netherlands). The signal was sampled online at 512 Hz and electrode impedances were kept below 15kΩ. The signal was referenced online to the Common Mode Sense (CMS) active electrode, which is placed in the center of all measuring electrodes and subtracted from the signal prior to offline processing.

The EEG signal was pre-processed using EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) toolboxes in MATLAB (The Math-Works, n.d.). The EEG signal was filtered offline using a bandpass filter of 0.1 – 30 Hz, and was referenced offline to an average of the two mastoids.

The first step of artifact rejection was performed using EEGLAB's infomax algorithm for independent component analysis (ICA), through which blink and horizontal eye movement components were manually identified and removed from the data. Between zero and three eye blink components were removed for all subjects (M = 1.6 components, SD = 1.13). Three of thirteen CI subjects had visible CI artifacts on electrode sites of interest that were removed using ICA. Electrode sites that were located over the CI transmitter magnet were eliminated from current analysis. These included sites, P7/8 and T7/8. These sites typically contained CI artifacts due to the fact that they were located at the sites of the CI transmitter magnet and we were often unable to establish a good connection between electrodes and scalp. F7 and F8 were also not included in the present analysis due to noise in these sites for many subjects (n = 4 CI; n = 7 TH).[6] The remaining electrodes were organized into regions that would be used in subsequent analyses: Occipital (O1, O2), Frontal (Fz, F3, F4), Central (Cz, C3, C4), and Parietal (Pz, P3, P4). The two Occipital electrodes were analyzed for the P1 component with a two-level factor of electrode, while the other electrodes in the Frontal, Central, and Parietal regions were entered as a factor of Region into the analyses for the other components (see Results section for more detail on which Regions were analyzed for each component).

The continuous signal was then segmented into 1200 ms epochs from -200 ms before to 1000 ms after stimulus onset. The second step of artifact rejection involved a voltage threshold of ±120 microvolts on channels of interest; all trials containing voltages over 120 microvolts were rejected. Remaining trials were visually inspected individually and deleted if any artifact remained. On average, each subject retained 90.1% of their trials after this process (range 44.3% - 100%). There was no difference in percent of trials rejected between CI and TH groups; *M*(TH) = 10.2%, *M*(CI) = 6.6%, *t*(33) = 0.95, *p* = .348.

ERPs were calculated for trials by condition for each subject, with the ERP time-locked to the onset of either the video (for visual P1) or the meaningful audio content within the audiovisual word target (for all other components). Each subjects' waveforms were visually inspected in order to determine the window during which components of interest were present for measurement purposes. For plotting group and condition differences, grand averages for both groups were produced.

[6] We identified this issue as faulty electrodes on two of the electrode looms in use, but had already collected a large proportion of data with these looms. Given that these sites are not as crucial for observing the ERP components of interest relative to the other electrodes, we elected to remove F7 and F8 from analyses.

Repeated Measures Analyses of Variance (rANOVA) was used to assess group and condition differences in behavioral measures, component amplitude, and peak latency. The specific between- and within-subjects factors used in each analysis are detailed below. Prior to computing the rANOVA, any extreme outlier values (as defined by values +/- 3 standard deviations from the mean) were excluded from analysis. Greenhouse-Geisser sphericity corrections were applied to the degrees of freedom of any variables where this assumption was violated. The rANOVA analyses were performed in R (R Core Team, 2023). Pearson correlation and partial correlation analyses were performed in Python using the pingouin-stats package (Vallat, 2018). An alpha value of .05 was considered significant in all analyses. Effect sizes in the form of partial eta squared are reported as well.

Note that for the P1 and N1 components, we initially planned to only include responses to match trials as a baseline comparison of sensory processes. Upon analysis, however, it was discovered that there was no difference in the pattern of results when the average of all trials was used, regardless of condition, instead of match-only trials. We therefore elected to use the average of all trials as this gives the results more statistical power. Furthermore, on early sensory components such as the P1 and N1, we decided to include all trials regardless of whether the subject responded correctly to the target. This was justified by the idea that early, obligatory sensory components should not be affected by later cognitive processes that would be associated with task accuracy. Additionally, filtering and ICA through the ERP pre-processing stage had already accounted for noise in the EEG data, eye-blink activity, and any CI artifact that might otherwise influence the measurement of early sensory ERP components.

For later components (N280/P300, N400, and Late N400), we first investigated condition differences within each group (CI and TH) using rANOVA, followed by planned post-hoc analyses to investigate any significant main effects or interactions between groups. Following any different patterns of responses across the two groups, we assess group differences using a rANOVA with a between-subjects factor of Group on the difference wave of each condition.

Results

Behavioral Results

Vocabulary. Percentile scores of the EOWPVT and ROWPVT were calculated based on a child's performance relative to other children of the same age. For receptive vocabulary (ROWPVT), the mean percentile for the CI-using children was 60.46 (*SD* = 32.21) and the mean percentile for the TH children was 83.18 (*SD* = 19.91). A t-test of the group differences indicated that the hearing group outperformed the CI group; $t(33) = 2.59, p < .05$.

On the measure of expressive vocabulary (EOWPVT), the mean percentile for the CI-using children was 54.23 (*SD* = 29.72) and the mean percentile for the TH children was 72.86 (*SD* = 21.54). A t-test of group differences revealed that the hearing group outperformed the CI group on the EOWPVT, $t(33) = 2.15, p < .05$.

Accuracy and Response Time. We calculated mean response time (RT) and accuracy on the picture-word matching task, both of which are presented in **Table 2** and visually-represented in **Figures 3a** and **3b**. We did not include trials in the behavioral analysis for response times more than 2.5 standard deviations below or above the mean for each condition. For the RT analysis, we only include values for trials on which children responded correctly.

Table 2. Mean reaction time (relative to target video onset) and mean percent accuracy for the picture-word matching task.

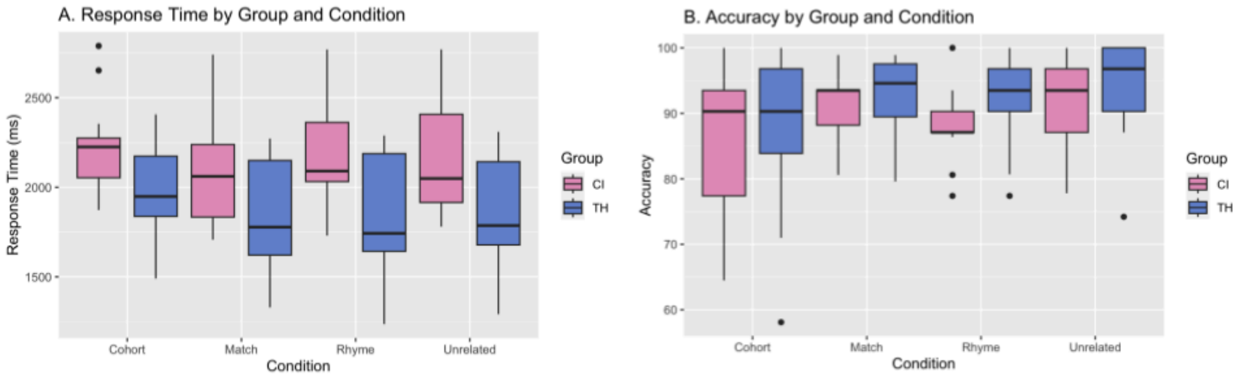
Condition	TH Group (n = 22)		CI Group (n = 13)	
	RT (ms)	Accuracy	RT (ms)	Accuracy
Cohort	1981 (*232*)	89.1 (*11.0*)	2217 (*265*)	86.2 (*10.1*)
Rhyme	1852 (*318*)	92.7 (*6.2*)	2179 (*276*)	88.5 (*5.7*)
Unrelated	1852 (*290*)	94.7 (*6.9*)	2138 (*298*)	91.1 (*6.8*)
Match	1835 (*305*)	92.4 (*5.6*)	2080 (*305*)	91.2 (*5.5*)

Note: values in parentheses represent standard deviations.

Both groups demonstrated similar levels of accuracy on the task. This was confirmed using a 2 (Group) x 4 (Condition) repeated-measures analyses of variance (ANOVAs) for accuracy, through which we observed an effect on Condition ($F(3, 99) = 5.15, p < .05, \eta_p^2 = .14$), with poorer accuracy for the cohort condition compared to the unrelated condition ($t(68) = 2.49$, $p < .05$). However, there was no main effect of Group ($F(1, 33) = 2.25, p = .14, \eta_p^2 = .06$), and no interaction; $F(3, 99)=0.40, p = .64, \eta_p^2 = .01$.

The rANOVA for RT between conditions and groups revealed an effect of Condition $F(3, 99) = 13.21, p < .0001, \eta_p^2 = .29$; as well as an effect of Group; $F(1, 33) = 8.07, p < .05, \eta_p^2 = .20$. There was no interaction; $F(3, 99) = 1.60, p = 0.195, \eta_p^2 = .05$. Through a post-hoc analysis to determine the nature of these main effects, we observed overall slower response times for the CI children ($t(33) = 2.66$, p < .001), but none of the differences between conditions were significant (all p's > .36).

Figure 3. Behavioral results



Figures 3a and 3b. Box-plots demonstrating the distribution of response times and task accuracy divided by Group and Condition.

ERP Results

Grand average waveforms depicting the visually-evoked ERP component P1 over occipital sites can be seen in **Figure 4**. In response to onset of the audiovisual word target, waveforms for both groups demonstrate a clear positive peak around 120 ms post-target stimulus for occipital sites O1 and O2 (P1).

Figures 5a and **5b** depict additional grand average waveforms in response to the audiovisual word targets, shown for the representative site Cz and separated by condition and group (for grand average waveforms in all electrode sites, refer to **Appendix B**). See **Figure 6**, where the baseline match condition for each group at Cz is plotted, for a more detailed view of where the measurement windows were for each of the components analyzed (for mean amplitude values at Cz organized by component measurement window and group, refer to **Appendix C**). These waveforms demonstrate apparent group differences. Both TH and CI groups show a prominent, widely-distributed negative peak between 150-200 ms (N1). In the TH group, the N1 is followed by a positive-going waveform from 200 - 350 ms (P300) maximally observed in central and parietal sites and a widespread persistent negative-going deflection beginning at 400 ms and continuing throughout the epoch (N400, Late N400). In the CI group, the N1 peak is followed by a series of centro-parietally distributed negative deflections from 200 - 350 (N280) and another widespread negativity beginning at 400 ms that continues late into the epoch (N400, Late N400).

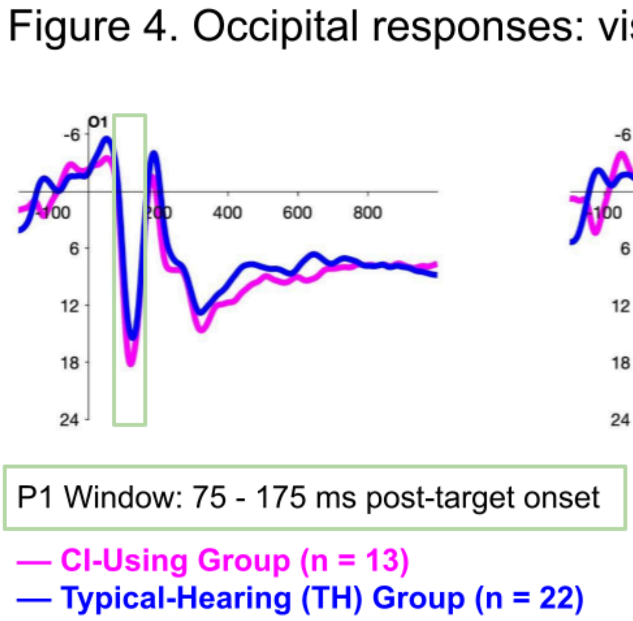


Figure 4. Grand average waveforms at occipital sites O1 and O2 depicting visual P1 responses, measured in the boxed area (75 - 175 ms after the onset of the target). The blue waveform shows the TH control group's visual responses and the magenta waveform shows the CI-using group's visual responses. When measuring P1 peak amplitude, we found no significant main effect of Group but did observe a main effect of Electrode, such that P1 responses at O2 were more positive than O1; ($t(67) = 2.49$, $p < .05$).

Visual P1 Amplitude. We measured Visual P1 peak amplitude in O1 and O2 from 75 - 175 ms after the onset of the audiovisual word target. This window was selected based on previous studies of the P1 and by inspecting the present waveforms (Corina et al., 2017; de Schonen et al., 2018).

The rANOVA for evaluating differences in P1 peak amplitude included the between-subjects effect of Group (TH or CI) and within-subjects effects of Electrode (O1 or O2). Our results show a main effect of Electrode; $F(1, 32) = 8.79$, $p < .05$, $\eta_p^2 = .22$. There was no main effect of Group ($F(1, 32) = 0.18$, $p = .674$, $\eta_p^2 = .01$) or significant interaction ($F(1, 32) = 1.86$, $p = .18$, $\eta_p^2 = .06$). A post-hoc t-test determined that within the main effect of Electrode, O2 had a more positive peak amplitude than O1 during the P1 window ($t(67) = 2.49$, $p < .05$). See **Table 3** for a summary of the ANOVA results for amplitude and latency measures of early sensory components (Visual P1 and Auditory N1).

Visual P1 Latency. Identical to the rANOVA for assessing peak amplitude differences, in the rANOVA for P1 peak latency we included the between-subjects effect of Group (TH or CI) and within-subjects effects of Electrode (O1 or O2). The results indicate no significant effects of Group ($F(1, 33) = 0.02$, $p = .877$, $\eta_p^2 < .00$), Electrode ($F(1, 33) = 3.89$, $p = .06$, $\eta_p^2 = .01$), or the Group*Electrode interaction on P1 peak latency ($F(1, 33) = 1.74$, $p = .20$, $\eta_p^2 = .05$).

Table 3. Summary of ANOVAs comparing the difference in peak amplitude and latency between groups on early sensory components (Visual P1, Auditory N1). This ANOVA had one between-subjects factor of group (CI, TH), and the within-subjects factor of Electrode (Visual P1: O1, O2) or Region (Auditory N1: Frontal, Central, Parietal). Degrees of freedom represent Greenhouse-Geiser corrections for sphericity violations, as well as removal of outliers that were more than 3 standard deviations away from the mean.

Effect	df (Group)	Group	df (Electrode *or* Region)	Electrode *or* Region	Group · Electrode *or* Region
		F		*F*	*F*
		p / η^2_p		p / η^2_p	p / η^2_p
Visual P1					
Peak Amplitude	1, 32	0.18	1, 32	**8.79**	1.86
		.67 / .01		**.01 / .22**	.18 / .06
Peak Latency	1, 33	.02	1, 33	3.89	1.74
		.88 / .001		.06 / .11	.20 / .05
Auditory N1					
Peak Amplitude	1, 33	0.003	1.28, 42.34	**10.49**	1.57
		.98 / .0001		**.001 / .24**	.22 / .05
Peak Latency	1, 33	1.70	1.32, 43.61	**83.43**	3.09
		.20 / .05		**<.000 / .72**	.08 / .09

Note: Bold text indicates significance.

Auditory N1 Amplitude. We measured Auditory N1 mean amplitude from 100 - 200 ms after the onset of the audiovisual word target, based on visual inspection and previous studies (Corina et al., 2022; Malins et al., 2013). As stated above, we opted to include all trials, regardless of condition, in the N1 analysis in order to increase statistical power. We performed a rANOVA for N1 peak amplitude with a between-subjects effect of Group (TH, CI) and within-subjects effect of Region (Frontal, Central, Parietal), plus the Group*Region interaction. We observed no significant main effect of Group ($F(1, 33) = 0.003$, $p = .96$, $\eta_p^2 < .00$) or interaction ($F(1.28, 42.34) = 1.57$, $p = .22$, $\eta_p^2 = .05$). There was a significant main effect of Region; $F(1.28, 42.34) = 10.49$, $p < .005$, $\eta_p^2 = .24$.

We conducted planned post-hoc pairwise t-tests to uncover the source of the Region main effect by analyzing differences between the Frontal, Central and Parietal regions above and beyond Group, but this analysis did not find any significant differences in N1 peak amplitude between regions upon applying Bonferroni corrections for multiple comparisons (all *p*'s > .16).

Auditory N1 Latency. For the rANOVA for N1 peak latency, we included the between-subjects effect of Group (TH or CI) and within-subjects effects of Region (Frontal, Central, Parietal), plus the Group*Region interaction. The results indicate a highly significant main effect of Region; $F(1.32, 43.61) = 83.43$, $p < .0001$, $\eta_p^2 = .72$. There was no significant main effect of Group ($F(1, 33) = 1.70$, $p = .20$, $\eta_p^2 = .05$) or interaction of Group with Region ($F(1.32, 43.61) = 3.09$, $p = .08$, $\eta_p^2 = .09$).

To follow up, we conducted Bonferroni-corrected pairwise t-tests at each region to determine the source of the region effect on N1 peak latency. The results of this post-hoc analysis suggest that there was no difference in N1 peak latency between frontal and central sites ($t(68) =$

1.44, *p* = .15), but that the latency of the N1 at parietal sites was significantly longer from both frontal (*t*(68) = 10.54, *p* < .0001) and central sites (*t*(68) = 8.85, *p* < .0001).

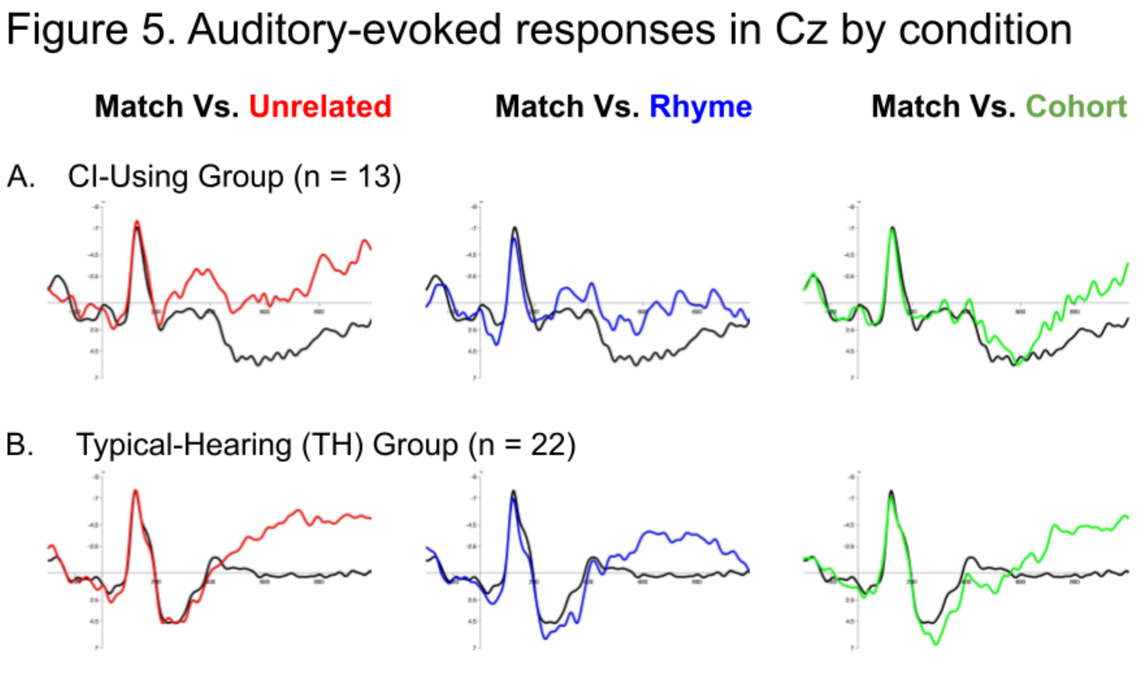


Figure 5. Grand average waveforms at Cz divided by Group and Condition, with the colored lines representing grand average response on the three different incongruent conditions (unrelated, rhyme, cohort), and the black lines representing the baseline match response for each group. Responses are time-locked to the beginning of the audio in the target word. Refer to Figure 6 for specific windows of measurement for each component.

Figure 6. Component time windows of measurement

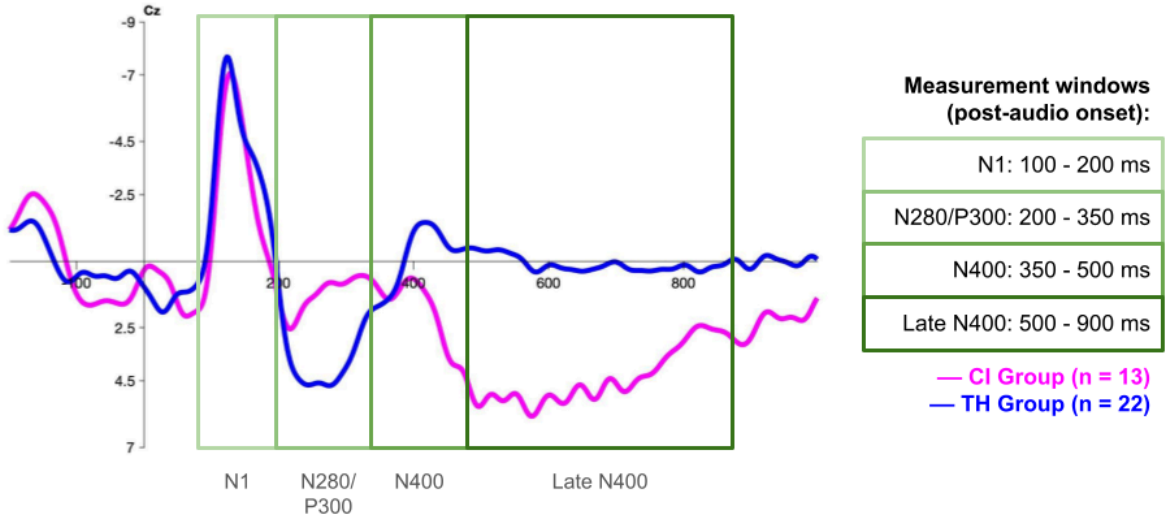


Figure 6. This grand average waveform taken from site Cz demonstrates the time windows used to take measurements of peak (N1) or mean (all other components) amplitudes. The specific electrodes and/or regions measured vary by component. The responses plotted here are the baseline (match) responses for the TH control group (blue) and the CI-using group (magenta).

N280/P300 Amplitude. As stated previously, in the window from 200 - 350 ms after target onset we observed clear group differences in grand average waveforms. The CI children demonstrated a centroparietal negative deflection that appears to be an N280 component, while TH controls show a centroparietal positivity resemblant of a P300. Therefore, we elected to categorize these groups and their components separately, despite using measurements taken in the same window and regions.

Cochlear Implant-Using Children: N280. We measured mean amplitude in the time window of 200 - 350 ms after the onset of the target and refer to this negative component as an N280. Measurements were taken in central and parietal electrode sites Cz, C3, C4, Pz, P3, and P4. A 4 (Condition) by 2 (Region: Central, Parietal) rANOVA did not reveal a significant main effect of

Condition ($F(3, 36) = 0.20$, $p = .89$, $\eta_p^2 = .02$) or interaction; $F(3, 36) = 1.03$, $p = .39$, $\eta_p^2 = .08$. A main effect of Region ($F(1, 12) = 19.30$, $p < .001$, $\eta_p^2 = .62$) demonstrated that the N280 was more negative in the central region than the parietal region; $t(102) = 4.37$, $p < .001$.

Typical Hearing Children: P300. In the window of 200 - 350 ms after the target onset, we took measurements of mean amplitude from the TH group for a P300 component in central and parietal sites Cz, C3, C4, Pz, P3, and P4. Using a rANOVA with within-subjects factors of Condition and Region, we observed no significant main effect of Condition ($F(2.21, 46.44) = 1.34$, $p = .27$, $\eta_p^2 = .06$) or interaction of Condition*Electrode; $F(2.13, 44.64) = 1.38$, $p = .26$, $\eta_p^2 = .06$. There was a highly significant main effect of Region; $F(1, 21) = 36.24$, $p < .0001$, $\eta_p^2 = .63$. Following up on the Region main effect, a t-test revealed that the P300 mean amplitude was more positive in the parietal region relative to the central region; $t(174) = 4.69$, $p < .001$.

N400 Amplitude. *Cochlear Implant-Using Children.* The mean amplitude for the N400 component was measured from 350 - 500 ms after the onset of the target in frontal, central, and parietal sites. This window was selected based on visual inspection, and happened to overlap with the well-established window typically used to measure the N400 component. Within the CI-using subjects, a rANOVA with within-subjects factors of Condition and Region demonstrated a significant main effect of Region ($F(2, 24) = 16.64$, $p < .0001$, $\eta_p^2 = .58$), but no effect of Condition ($F(3, 36) = 0.75$, $p = .53$, $\eta_p^2 = .06$) or significant interaction of Condition*Region; $F(2.98, 35.76) = 1.25$, $p = .31$, $\eta_p^2 = .09$. Bonferroni-corrected pairwise t-tests on the mean N400 amplitude at each region, regardless of condition, suggested that the amplitude of the N400 over parietal sites was significantly less negative from frontal ($t(102) = 4.71$, $p < .0001$) and central regions ($t(102) = 4.72$, $p < .0001$), but that frontal and central regions did not significantly differ from each other in N400 amplitude; $t(102) = 0.49$, $p = .62$.

Typical Hearing Children. A rANOVA was performed on the N400 amplitude of the TH group, with within-subjects factors of Condition and Region. This analysis showed a significant main effect of Region ($F(1.23, 25.80) = 19.98$, $p < .0001$, $\eta_p^2 = .49$), but we did not observe a main effect of Condition ($F(3, 63) = 1.07$, $p = .37$, $\eta_p^2 = .05$) or significant interaction ($F(3.09, 64.94) = 1.04$, $p = .38$, $\eta_p^2 = .05$) on N400 mean amplitude in this window. Post-hoc t-tests on the difference in amplitudes between regions suggest, similar to the CI-using group, that N400 mean amplitude was significantly less negative in the parietal region from the other two regions; Frontal/Parietal: $t(174) = 6.38$, $p < .0001$; Central/Parietal: $t(174) = 6.23$, $p < .0001$; Frontal/Central: $t(174) = 0.71$, $p = .48$.

Late N400 Amplitude. *Cochlear Implant-Using Children.* We measured the mean amplitude of the “Late N400” component from 500 - 900 ms after the onset of the target, using the same frontal, central, and parietal regions as the N400. This window was selected based on visual inspection of both groups’ waveforms. The rANOVA for Late N400 mean amplitude with the CI group demonstrated significant main effects of Condition ($F(3, 36) = 4.88$, $p < .01$, $\eta_p^2 = .29$) and Region; $F(1.27, 15.19) = 12.16$, $p < .01$, $\eta_p^2 = .50$. There was also a significant interaction of Condition*Region; $F(6, 72) = 2.30$, $p < .05$, $\eta_p^2 = .16$. Post-hoc analyses comparing conditions at each region were conducted to explore the nature of this interaction. This analysis indicates that there were condition differences in Late N400 mean amplitudes in the central region ($F(3, 36) = 5.11$, $p < .05$, $\eta_p^2 = .30$) and parietal region ($F(3, 36) = 6.00$, $p < .05$, $\eta_p^2 = .33$), but not in the frontal region; $F(3, 36) = 2.72$, $p > .05$, $\eta_p^2 = .18$. Further Bonferroni-corrected pairwise t-tests to assess condition differences within central and parietal regions did not show any significant mean differences in Late N400 amplitude (all p’s > .15).

Typical Hearing Children. The rANOVA on the Late N400 mean amplitude for the TH children produced a significant main effect of Condition; $F(3, 63) = 5.75, p < .001, \eta_p^2 = .22$. Pairwise t-tests between each condition, averaged across regions and adjusted for multiple comparisons, did not reveal statistically significant differences between mean amplitude across conditions. Some of the condition differences, however, were trending towards being significantly different: the Unrelated condition had a more negative Late N400 amplitude during the Late N400 window than the Match condition ($t(130) = 2.60, p = .051$). The Unrelated condition also had a slightly more negative N400 amplitude in this window relative to the Cohort condition; ($t(130) = 2.47, p = .061$). Other condition differences were not significant (all *p*'s > .19).

Furthermore, there was a significant main effect of Region ($F(1.16, 24.35) = 20.57, p < .001, \eta_p^2 = .50$), where the N400 amplitude in the parietal region was significantly more positive than the frontal region ($t(174) = 7.37, p < .001$) and central region; $t(174) = 5.61, p < .001$. There was no difference between N400 amplitude in the frontal and central regions; $t(174) = 2.17, p > .05$.

Groupwise Comparisons. Based on the rANOVA results within groups and visual inspection of the two groups' grand average waveforms, we performed a between-groups comparison of the differences between the different mismatch conditions (Unrelated, Rhyme, Cohort) and the baseline Match condition. This subtraction method allows us to draw inferences on the "effect size" of the manipulations of semantic congruency on components like the N400 and Late N400. Refer to **Figure 7** for difference waves for each group plotted by mismatch condition at site Cz. The following analyses were performed using a between-subjects factor of Group (CI or TH), within-subjects factors of Condition (Unrelated, Rhyme, Cohort - all with Match subtracted) and

Region (Frontal, Central, Parietal). We also included the interaction terms for these factors. The results of these ANOVAs for the N400 and Late N400 are presented in **Table 4**.

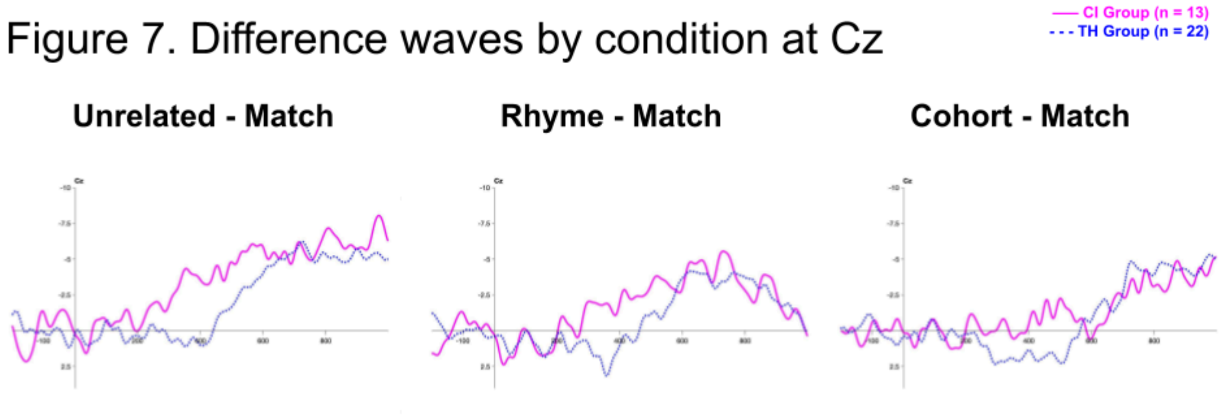


Figure 7. Difference waves by mismatch condition and Group at site Cz.

N400. In the window of 350 - 500 ms after the target onset, we did not observe any significant main effects of Group ($F(1, 33) = 3.68$, $p = .06$, $\eta_p^2 = .10$) or Condition ($F(2, 66) = 0.82$, $p = .44$, $\eta_p^2 = .02$) on subtraction wave amplitudes. There was a main effect of Region; $F(1.36, 44.73) = 7.15$, $p < .05$, $\eta_p^2 = .18$. Post-hoc Bonferroni-corrected pairwise t-tests of region differences did not produce any significant differences between subtraction waveform amplitudes at the three regions (all *p*'s > .05). We did not find any significant interactions between any of these factors (all *p*'s > .65).

Late N400. The analysis of difference waveforms from 500 - 900 ms after target onset revealed no significant main effect of Group ($F(1, 33) = 0.15$, $p = .70$, $\eta_p^2 < .00$) or Condition; $F(2, 66) = 0.99$, $p = .38$, $\eta_p^2 = .03$. We observed a significant main effect of Region; $F(1.67, 55.27) = 11.05$, $p < .001$, $\eta_p^2 = .25$. Post-hoc pairwise t-tests on each region suggest that there were significant differences in the subtraction waveforms (regardless of group or condition) between frontal and the other regions; Frontal/Central: $t(208) = 2.90$, $p < .05$; Frontal/Parietal: $t(208) = 3.96$, $p < .05$; Central/Parietal: $t(208) = 1.08$, $p = .28$. For the Late N400 difference waves, we did not find any significant interactions (all *p*'s > .14).

Table 4. Summary of groupwise ANOVAs comparing the difference in amplitude between match and other conditions during the windows of the N400 and Late N400. These ANOVAs had one between-subjects factor of group (CI, TH), and within-subjects factors of Condition Difference (Rhyme, Unrelated, Cohort) and Region (Fronal, Central, Parietal).

Effect	N400 df	N400	Late N400 df	Late N400
Group *F*	1, 33	3.68	1, 33	0.15
p / η^2_p		.06 / .10		.70 / .004
Condition *F*	2, 66	0.82	2, 66	0.99
p / η^2_p		.44 / .02		.38 / .03
Region *F*	1.36, 44.73	**7.15**	1.67, 55.27	**11.05**
p / η^2_p		**.006 / .18**		**<.000 / .25**
Group · Condition *F*	2, 66	0.15	2, 66	0.73
p / η^2_p		.86 / .01		.49 / .02
Group · Region *F*	1.36, 44.73	0.25	1.67, 55.27	0.11
p / η^2_p		.69 / .01		.87 / .003
Condition · Region *F*	2.52, 83	0.49	2.55, 84.09	1.95
p / η^2_p		.66 / .02		.14 / .06
Group · Region · Condition *F*	2.52, 83	0.31	2.55, 84.09	1.16
p / η^2_p		.78 / .01		.33 / .03

Note: Bold text indicates significance.

Correlation Analyses

We performed Pearson correlation (and where applicable, partial correlation) analyses between ERP measures and specific demographic, behavioral, or language related variables of interest. The results of these correlation analyses are reported in **Appendices D-F**.

Age and CI hearing factors. First we performed correlations between subjects' chronological age and the peak amplitude and latency of their early sensory components, the visual P1 and the auditory N1. We observed no associations between peak amplitude and age for either component, in either group (all *p*'s > .17). The TH subjects showed negative correlations between component latency and age; P1 Latency and Age: *r* = -.46, *p* < .05; N1 Latency and Age: *r* = -.44, *p* < .05. We did not observe this same association for the CI using subjects; P1 Latency and Age: *r* = .46, *p* = .11; N1 Latency and Age: *r* = -.02, *p* = .96. Scatter plots showing these associations by group are depicted in **Figure 8**.

Next, we tested for associations between CI users' Time in Sound (TIS) and Age of Implantation (AOI) with early sensory component peak amplitude and latency. For these factors, we opted to run a partial correlation with Chronological Age as a covariate, in order to determine any associations between CI hearing experience and P1 or N1 outcomes, above and beyond the effects of age. We did not observe any significant correlations between factors of our CI using subjects' hearing experience and visual P1 peak amplitude or latency (all *p*'s > .71). There was also no significant association between TIS or AOI with N1 peak amplitude or latency (all *p*'s > .15).

Behavioral task performance. We tested for associations between task performance (accuracy and response time) and either N280 amplitude or P300 amplitude, depending on subjects' groups. For TH children, we calculated a single P300 mean amplitude as a weighted average of all four conditions, from all six centroparietal electrodes where this component was observed. This approach was used because the P300 is not suspected to be affected by condition differences. We correlated this measure with overall task accuracy and RT on trials on correct trials. We did not observe a strong association between P300 mean amplitude and either of these measures of task performance (all *p*'s > .14).

For the CI using children, we decided to test for task performance associations with both the baseline N280 MNA response (calculated from all six centroparietal sites where we measured the N280), as well as the difference between this baseline response and each incongruent condition response in the same window. We observed a strong positive correlation between N280 baseline mean amplitude and overall task accuracy; *r* = .68, *p* < .05. A moderate negative correlation between the Cohort N480 difference value and overall RT was also discovered; *r* = -.55, *p* = .05.

Language measures. We ran correlation analyses using standardized percentile scores for the Receptive and Expressive One Word Picture Vocabulary Tests (ROWPVT and EOWPVT, respectively); these standardized scores account for changes in vocabulary size that happen with Chronological Age. We tested for correlations between the ROWPVT/EOWPVT and the effect size of the N400 and Late N400. Effect size was calculated by subtracting the baseline "Match" condition amplitude from the incongruous condition amplitudes for each subject in both the N400 and Late N400 windows. The greater the difference between semantically congruent and incongruent responses, the greater the "N400 effect size". We tested for associations between vocabulary measures and condition N400 effect size (e.g., averaging the effect size values for each condition across the three regions). This was done for both groups; we did not observe significant correlations for either group (CI: all *p*'s > .22; TH: all *p*'s > .24).

Figure 8. Correlations of sensory component latency with chronological age

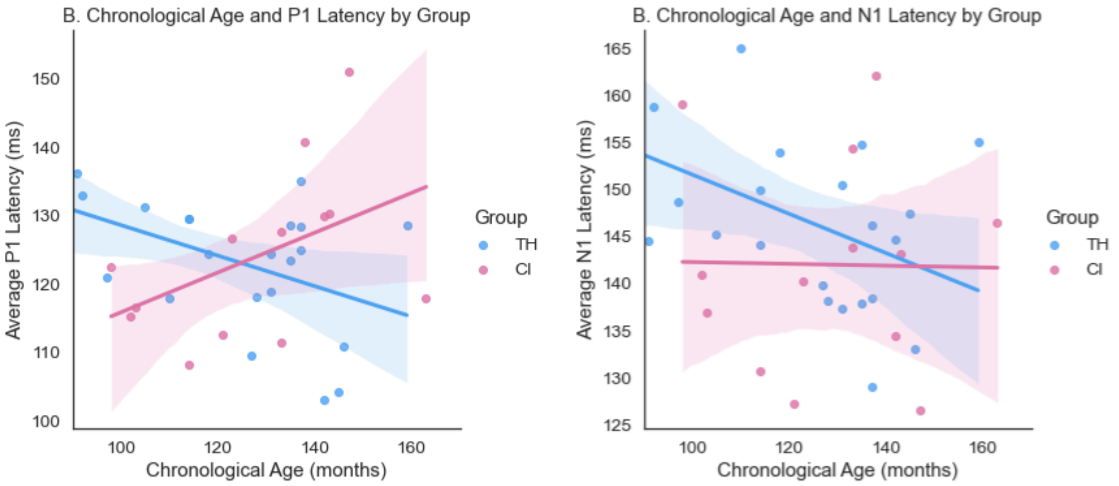


Figure 8. Pearson correlations are depicted as scatter plots with 95% confidence intervals, separated by sensory components (P1, N1) and group. These scatter plots show the association between component latency and chronological age of the subjects. For both components, the TH subjects showed negative correlations between component latency and age; P1 Latency and Age: r = -.46, p < .05; N1 Latency and Age: r = -.44, p < .05. We did not observe this same association for the CI using subjects; P1 Latency and Age: r = .46, p = .11; N1 Latency and Age: r = -.02, p = .96.

Discussion

The present study investigates the neural correlates of naturalistic audiovisual spoken word processing in children with typical hearing and those who hear through cochlear implants. Through manipulation of phonological and semantic congruity of picture primes and audiovisual word targets, we aimed to detect group differences in the neural mechanisms that support the processing of naturalistic aspects of spoken words. Further, this study investigated the association between language, hearing, and age-related factors with neural responses during audiovisual spoken word recognition. The results indicate that while the CI-using group was overall slower to respond on the picture-word priming task and underperformed relative to TH controls on measures of vocabulary, there were very few differences in neural indices of their ability to recognize audiovisual spoken words during the picture-word priming task. We discuss results specific to each ERP component below.

It was initially hypothesized that the magnitude of the P1 response for CI-using children may be larger than that of children with typical-hearing, under the assumption that visual speech information would be more salient for CI-users who might have less reliable auditory input. CI-using children could have also shown faster P1 latencies than TH children. These patterns of results would be consistent with other studies comparing visually-evoked potentials between TH and CI-using children (Corina et al., 2024; Pierotti et al., *in preparation*). The present findings do not suggest that there is a difference between groups in the magnitude or timing of visually-evoked neural responses to audiovisual speech stimuli.[7] We also did not observe any associations above and beyond chronological age between CI hearing characteristics (e.g., AOI or TIS) and P1 amplitude or latency. These findings are surprising given that the stimuli used in both Pierotti et al. (*in preparation*) and the current study are similar: both studies used short AV word targets presented with pictures. That said, a lack of P1 amplitude and latency differences between TH and CI-using children has been reported previously, and the current findings are aligned with these studies (Corina et al., 2017; Liu et al., 2017).

[7] See **Appendix G** for the grand average waveforms for each groups' responses to the static speaker images. We only recorded six of these events per subject, and therefore did not have enough data to formally analyze these responses. Upon visual inspection of the waveforms, however, we note the presence of typical visual componentry such as P1 responses and a N170, which is a well-known ERP component associated with face processing. Visual inspection does not suggest that there are significant group differences between responses to the static speaker images.

It should be noted that the present study uses a much more narrow and older age range of CI-using subjects than other studies investigating the P1 in this population, and that the subjects in the present study had less variability in their age of implantation and duration of CI use - both factors that could be associated with visual speech cue processing (Corina et al., 2024; Schorr et al., 2005). While we refrain from overinterpreting this null finding, it is possible that these differences in CI subject characteristics contribute to the different outcomes we observe between studies. More specifically, the children in our study may not have developed a reliance upon visual speech cues in the same way as children who receive their implant relatively late into childhood. Furthermore, we acknowledge that we did not include audio-only and video-only (lip-reading) trials. This would have given us greater insight into the unique patterns of neural responses attributed to each sensory modality. The decision not to include these sensory modality manipulations was deliberate because the duration of the task was already long enough for children of this age group. However, in future extensions it would be possible to run children on different versions of the same experiment or to have the same children complete different versions across multiple sessions.

Next, regarding the N1, we hypothesized that the latency and amplitude of this auditory-evoked component would be smaller (less negative) and slower to onset in CI-users compared to TH children, as evidence of CI-users' less mature auditory processing systems and consistent with previous studies (Corina et al., 2017, 2022). At present, our results do not suggest that the CI-using children in our study differed from typical hearing controls in their early auditory processing as indexed by the amplitude and latency of the N1 component. Additionally, no significant correlations were observed between TIS or AOI and N1 amplitude or latency, controlling for chronological age. We observed that N1 latencies were longer in parietal than fronto-central sites, and directionally this effect may have been stronger for TH children than CI users, but the difference was not statistically significant and perhaps not as meaningful given that the N1 is consistently maximally observed in frontal sites.

Without overinterpreting the null finding of no group differences, we consider that the specific CI-using subjects in our study have had more time and experience with sound to develop mature auditory processing systems, relative to younger CI-users or those who had only had their CI for a few months as reported in other studies. All but one of the children in our study had been using CIs for over five years (the one other subject had used hearing aids for five years prior to receiving their CI). This duration of experience with sound, albeit different from naturalistic hearing, presumably allows children to develop auditory processing systems that could contribute to typical-looking early sensory ERP components. Similar to the visual P1 component, it is possible that our sample does not reflect the great variability of sound experience that is present in the CI-using child population, and therefore we are limited in our ability to observe group differences between our CI-using sample and the TH control group.

Interestingly, we observed group differences in associations between P1 and N1 peak latency and chronological age. Both P1 and N1 peak latencies seem to shorten as typical hearing children get older, but we did not see the same trends in the CI-using children in our study. Due to the cross-sectional nature of the present research, we can not draw conclusions about how early visual and auditory processing components change as a child develops. The direction of this association, however, is aligned with the idea of maturation and increasingly efficient processing of sensory systems as typical hearing children develop (Corina et al., 2022; Kaganovich et al., 2016; Knowland et al., 2014; Sharma et al., 2005). That this trend is seen in both auditory and visual modalities may also serve as evidence that these two systems are becoming efficient together, perhaps aligned with the U-shaped trajectory of AV speech processing (Jerger et al., 2009). At this time, we can not make a similar inference for the CI-using subjects in our experiment.

The present study aimed to elucidate any group differences in phonological processing or strategy used to complete this task by analyzing the N280 and P300 components. We predicted that the presence of an N280 component would reflect the mapping of speech sounds, with an "N280 effect" denoting more resources spent reconciling incongruent speech sounds in rhyme and unrelated word targets (Desroches et al., 2009, 2013; Malins et al., 2013). If phonological processing is especially challenging for CI users, or if these children are predictively pre-activating phonological information given that their bottom-up sensory input is less reliable, we anticipated that N280 effects could be larger than those of children with typical hearing. On the other hand, we anticipated that children with typical hearing may not elicit an N280 component during this task if they are utilizing a global task-level strategy of predicting trial congruency (e.g., categorization of if a trial is going to be a match or mismatch). In this case, a parietally-localized P300 component (e.g., a P3b) may be more likely to appear (Van Petten & Luka, 2012).

The present findings are partially aligned with the expectations described above: our CI subjects evoked a negative-going component that was maximally observed in the central region (N280), though its amplitude was not sensitive to phonological incongruity to the extent we anticipated (e.g., we did not observe a main effect of Condition). The TH group, however, elicited a parietal, positive, peak-like component that we interpreted as a P300 component.

We consider how the presence of an N280 versus a P300 may suggest the use of different strategies between CI-using and TH children in the context of the present task. As demonstrated through relatively high and equivalent accuracy across groups, this task was not particularly challenging for 8 to 13 year old children regardless of hearing status. We did observe group differences in response time, such that CI-using children were overall slower regardless of condition than TH controls. This could suggest that CI-using children may have been more deliberate or cautious in their responses; nonetheless, they still reached the correct response as consistently as typical hearing children. We may infer that CI-using children were devoting more time to processing specific speech sounds of the target words in a way that reflects a typical N280 response. The presence of the visual speech cues may have even encouraged or aided CI-using children as they attended to the words. It is unclear why there was not an effect of condition on the amplitude of the N280 for CI-using children, especially since it has been shown that AV speech can exaggerate condition differences in this window of processing (Brunellière et al., 2020), but it is possible that the phonological differences between matches and the other conditions were perceived as subtle enough by the CI-using children that the information was processed relatively equivalently.

To this point, there is evidence from research on children's ability to recognize speech in noise that suggests the selective attention skills needed to be able to isolate relevant speech cues may take extensive listening experience to develop (see Leibold & Buss, 2019 for a review of this literature). A notable concept within this literature is the idea of *perceptual weighting shift* (Nittrouer et al., 1993; Nittrouer, 2002): in typical hearing developmental populations, young children who are learning the relevant cues of their native language may attend more to dynamic acoustic speech signals (e.g., formant transitions), whereas children starting around age 7 and adults focus more on stable speech signals (e.g., frication noise). The development of this selective attention, especially in noisy contexts, is presumed to take time and listening experience to mature. It is possible that the CI users in our study had not yet fully developed this sensitivity to the stable speech cues in spoken language, and combined with unreliable bottom-up input through their CI, these children may have inefficiently processed the phonological information of each AV spoken word in an equivalent manner regardless of phonological congruity with the picture prime. Additionally, we observed an association between overall task accuracy and baseline (match) N280 mean amplitude, such that the CI-using children who were more accurate demonstrated less negative N280 baseline responses. Taken in this context, it is possible that within our CI-using subjects, those who did not need to allocate as many resources towards phonological processing (e.g., had less negative N280 responses) may have also had an easier time making a correct response during the task. These CI-users may have learned to attend to the relevant cues in their native spoken language, or perhaps unconsciously understood that predictive pre-activation of upcoming phonological information was futile when half of the trials in the task contained incongruent speech sounds from the picture prime. Both of these outcomes could be reflected in their neural responses and behavioral performance.

Shifting our discussion to the control group, TH children may have found the presence of both the audio and visual speech cues in this task to be redundant, making an already straightforward task that much easier. We presume audiovisual speech processing of single words in a distraction-free lab environment is rather effortless for the children in this study with typical hearing, especially considering that children ages 8 to 13 may have already unconsciously realized the benefit to utilizing both auditory and visual speech cues (see earlier discussion of the U-shaped trajectory of AV speech processing; Jerger et al., 2009; Lalonde & Holt, 2015).

Therefore these TH children may have employed a strategy that would benefit their goal of finishing the task as quickly and accurately as possible. Such a strategy may have been to attend to more global cues that require less cognitive load, including response probability, as opposed to specific within-trial cues such as congruency of phonological or semantic information between the prime and target. Rather than predictively pre-activate upcoming phonological information about a target, children in the TH group may have predicted whether each trial fell into either a "match" or "mismatch" category, as was crucial in the present study for being able to make a swift and accurate response. A requirement to overtly classify stimuli into one category or another has been shown to increase the amplitude of the P300 (Johnson, 1988). The presence of a P300 response, either in place of or overshadowing the presence of an N280 response, could reflect the disconfirmation of an expectation of one trial category or another (Van Petten & Luka, 2012). If the present task had been passive in nature (e.g., subjects had not been required to respond at the end of each trial), it is possible that the P300 response would not be as apparent.

The P300 component's amplitude was also not sensitive to the trial condition. This lack of condition sensitivity makes sense if the P300 reflects children's ability to categorize whether or not the target was simply a match or mismatch, as opposed to detecting more sensitive properties of the stimulus such as its phonological or semantic properties. In the current study, we had 50% match trials and 50% mismatch/incongruent trials; this might suggest that the P300 was evoked equally for either type of target, depending on whether a subject was predicting one type of target or the other. In a future extension of this work, it would be useful to manipulate the likelihood of match trials within the task in order to elucidate if the amplitude of the P300 was affected by the probability of a match trial. More specifically, given that the amplitude of the P300 component is sensitive to category probabilities, we may anticipate that if only 25% of trials were matches, we would only see P300 responses on those trials, with a larger amplitude relative to a version of the experiment with a greater probability of match trials (Folstein & Van Petten, 2011; Kutas et al., 1977).

It may be expected that the presence of a P300 would be associated with subject characteristics such as task performance or vocabulary scores. In other words, we might predict that the children with the largest vocabularies would find this task to be the easiest, and these children would be the ones to employ this more efficient, global strategy. Children's vocabulary (specifically receptive vocabulary) has been linked with greater performance on speech recognition tasks and mature listening strategies (McCreery & Stelmachowicz, 2011; although Eisenberg and colleagues, 2000, did not observe this relationship). In the present study, our correlation analysis of task accuracy, response time, and both receptive and expressive vocabulary did not yield strong associations with P300 amplitude in the 200 - 350 ms window. Therefore, we can not infer from the present group of typical hearing subjects that being a "strong performer" on this task or having a larger vocabulary is associated with the presence of a P300 response.

We consider why CI-using children in the present study did not elicit a P300 response like TH children, above and beyond strategy differences while completing the task. Recall that in a review from Beynon and Snik (2004) on the P300 in CI-users during speech recognition, it was discussed that CI-using children have demonstrated auditory P300 responses to pure tones but not to speech sounds on a simple discrimination task. While the present task differs in nature from simple syllable discrimination, it may still be true that if the auditory P300 is associated with acoustic and phonological processing, this could impact the manifestation (or lack thereof) of a P300 response for this group. We note that the lack of group differences in N1 responses as discussed previously might be taken as evidence that basic sensory processing of auditory information is not the cause of the P300 absence in CI users. Phonological processing differences, however, could be associated with the P300. As stated previously, it is possible that CI using children do not have as sensitive phonological processing capabilities as children with typical hearing, or they may have less mature selective attention abilities for filtering irrelevant speech signals. This could result in both an N280 response that does not modulate with phonological incongruence as well as a lack of a P300 response that may index the categorization of task-relevant speech information. However, we suggest that more work is needed to uncover the specific factors that may be associated with the presence of a P300 response as opposed to an N280 response in picture-word priming tasks such as this study. This is especially important because, to our knowledge, this is the first study to investigate the N280 component in CI-using children and to directly compare the presence of an N280 and a P300 component in the same time window.

With respect to the N400 component, we entertained several predictions. First, we expected condition differences such the trials with unrelated pairs of picture primes and audiovisual word targets would elicit more negative N400 amplitudes than those with matching primes and targets, suggesting greater processing costs associated with incongruent semantic content canonically known as the N400 effect (Lau et al., 2008). We further expected that rhymes may show an N400 effect, albeit not as strong as unrelated trials given that rhymes may receive spreading activation from expected target words. This spreading activation may lower their processing costs relative to completely semantically unrelated words (Praamstra & Stegeman, 1993). Word initial cohorts, on the other hand, may elicit longer latencies of the N400 relative to other incongruent conditions, given a later point of disambiguation between matches and cohorts that share phonological onsets (Connolly & Phillips, 1994; Desroches et al., 2009). As for group differences, we expected exaggerated unrelated N400 effect sizes for CI-using children compared to TH children based on previous research with this population that suggests CI-using children have a harder time reconciling unexpected semantic content (Kallioinen et al., 2016; Pierotti et al., 2021). Rhymes may be indistinguishable from match words if the phonological differences between these conditions are undetectable by CI-users. Word initial cohorts may take longer to resolve for CI-users, or similarly may be undetectable by this group.

The results of this study do not exactly confirm the predictions above. Surprisingly, we found no evidence of a condition difference (e.g., the classic N400 effect) for either group in the traditional N400 window of 350-500 ms after word onset. This specific finding is inconsistent with previous studies that have investigated the N400 in this population using similar stimuli. For example, Pierotti and colleagues (2021) used audiovisual word primes and picture targets to study the N400 in typical hearing and CI-using children ages 2 - 10 years, and observed robust N400 effects for both groups. We acknowledge that the present study differs from Pierotti et al. (2021) in that we use audiovisual spoken words as targets instead of primes. However, previous studies using AV targets have been able to demonstrate an N400 effect in both adult and developmental populations (Brunellière et al., 2020; Kaganovich et al., 2016), so this is unlikely to be a factor in why the main effect of condition was not statistically significant.

Though absent in the typical N400 window, effects of condition that aligned with the original hypotheses for this study were more apparent in the Late N400 window from 500 - 900 ms after the target onset. For the CI-using group, we observed an interaction of condition and region where condition differences emerged in central and parietal regions, but not in the frontal region. However, upon applying Bonferroni corrections to adjust for multiple comparisons, the differences between conditions were not statistically significant. Similarly, TH group demonstrated a main effect of condition, but condition differences were also not robust enough to be statistically significant.

The directionality of the effects in the TH group, however, aligns with our predictions of a semantic incongruity effect such that we observed more negative N400 amplitudes for the unrelated condition relative to the match and cohort conditions. Given the proposed graded nature of the N400 effect, it would make sense that the strongest (although not statistically significant) difference in N400 amplitude between conditions would occur for the target types that were the least overlapping in semantic content (Hendrickson et al., 2019; Kuperberg & Jaeger, 2016). Rhymes, which share some phonological information with the target and could benefit from spreading activation from the prime picture label, were not hypothesized to be as difficult to integrate as unrelated words for either CI-using or TH children. That said, it is interesting that we did not observe any condition- or region-level sensitivity to the rhymes in our N400 results, such as the frontally-localized late negativity that has been shown to be sensitive to rhyming spoken word stimuli in children as young as preschool ages (Andersson et al., 2018; Coch, Grossi, et al., 2002). We did not explicitly draw subjects' attention to the rhymes, or other phonological manipulations in our experiment. Therefore, it is possible that the location and magnitude of the N400 effect in the present experiment was buffered from the influence of rhyme targets' phonological overlap with the match targets.

It is surprising that the cohort effect was not statistically significant even as late into the epoch as 900 ms. Visual inspection of the waveforms suggests that perhaps our measurement window should have been extended even later into the epoch, however there were concerns about taking measurements in windows where there may be overlap with subjects' overt behavioral responses. The waveforms alone also do not clearly convey the variability in subjects' neural responses this late into the epoch, likely due to inconsistencies in the timing of the meaningful audio in our stimuli as opposed to true within- and between-subjects variability in component amplitudes. Later in this discussion, we speculate how stimuli timing issues could have contributed to the lack of significant results seen particularly in the N400/Late N400 windows.

Our final prediction regarding the N400/Late N400 was that the magnitude of the semantic incongruity effect measured in these windows may be correlated with subjects' vocabulary size. We anticipated that children with greater vocabulary scores may be more adept at semantic retrieval which could manifest as smaller differences between neural responses to semantically congruent and incongruent targets (see Schneider et al., 2023, for more on how semantic retrieval may be indexed by both the N400 and theta-band oscillatory responses, and the possible link of vocabulary to these processes). In the present study, however, we did not observe a statistically significant correlation between either receptive or expressive vocabulary and N400 effect size; this was true of subjects in both the CI-using and the typical hearing groups. The link between the N400 and vocabulary is tenuous, and other studies have also not been able to show an association between the two (Henderson et al., 2011; Pierotti et al., 2021). As discussed by Schneider and colleagues (2023), it is possible that the association between the N400 effect and vocabulary is specifically localized to the central left hemisphere; this granularity of localization was outside of the capabilities of the present study given the limited number of electrodes we employed, and the implicit challenges of measuring lateral electrodes with respect to the placement of CI transmitter magnets.

Aside from differences in N280/P300 components, the present study revealed few group differences between CI-using children and TH controls in patterns of neural responses supporting audiovisual spoken word recognition. We did, however, observe some differences in behavioral performance both on vocabulary measures and the task itself. As discussed, CI-using children were overall slower to respond to trials compared to TH children, but completed the task with commensurate levels of accuracy. CI-using children, as a whole, also performed worse on measures of expressive and receptive vocabulary compared to TH controls. In brief, while behavioral disparities emerged, it is not clear that these are due to differences in processing at a neural level. A similar finding was reported by Bell et al. (2019), who found that while CI-using underperformed on various language measures and a picture-word matching task, the CI-using children showed similar-looking N400 effects to typical hearing counterparts. Taken together, these studies could provide evidence that the source of differences in spoken word recognition outcomes is seemingly separable from neurophysiological responses that support these functions. However, it is also possible that group differences emerge during downstream decision-making processes, and that these are not reflected in the present study's waveforms. We refrain from further interpretation of the null findings in our study, given that the subjects in our CI-using group may not fully represent the diverse backgrounds, etiologies, linguistic experiences, and outcomes associated with this population.

Much of this study's design was based on the research from Desroches and colleagues (2009) and its extensions with developmental populations (Desroches et al., 2013; Malins et al., 2013). Therefore, it is worth discussing the potential source of differences in results observed between the present research and these original studies. First, while aiming to use the same age range of subjects as Desroches et al. (2013) and Malins et al. (2013) of children 8 to 12 years, we opted to broaden this range to up to 13 years so we could include more CI-using subjects' data. It is unlikely that this change had a great influence on the results we observed.

Additionally, while we used almost all of the same concept pairs and picture primes as these previous studies, we had to recreate our own versions of the word targets so that we could make use of both audio and visual speech cues. It is possible that the differences in word timing, pronunciation, prosody, and many other speaker-related factors contributed to the present pattern of results. Indeed, one of the key research questions addressed by Desroches et al., (2009) and its extensions is the temporal nature of the spoken word recognition process. The present research seems to provide evidence for how remarkably sensitive the neural markers of this process might be to the timing of audio speech cues. The subtle timing differences between the original study and the present study make it challenging to infer the true impact of the addition of visual speech cues (e.g., the video itself) on our results.

Working from the foundations of predictive coding and neurocognitive accounts, we initially theorized that visual speech cues would make contact with auditory speech cues early on in the process of spoken word recognition, in multisensory integration areas localized in the sensorimotor cortex. Downstream predictions formed about both phonological and lexical-semantic information may therefore be influenced by the information provided by the visual modality (such as visemes). This could be especially true for CI-using children, who are faced with unreliable bottom-up input in the auditory domain, and may be reflected in stronger evidence of predictive processing of phonological and semantic content from this group relative to TH controls. We reconcile the present results with this theory. On one hand, our study was not designed to be able to isolate the region or location where audio and visual speech cues are integrated. Imaging methods with greater spatial resolution, such as magnetic resonance imaging (MRI) or magnetoencephalography (MEG), would be more useful than EEG in targeting specific brain regions of interest. Further, we are unable to identify the unique contributions of both audio and visual cues to children's understanding of a spoken word. In order to disentangle these two modalities, we will need to collect additional neural responses from audio-only and visual-only trials.

On the other hand, we do observe group differences in the waveform morphology occurring in the 200 - 350 ms window after the word onset, which could reflect prediction strategy differences as discussed previously. While the CI-using children showed an N280 response that may suggest they are forming predictions about phonological information in the AV word target based on cues from the picture prime, we observed a P300-like response in the TH controls that is suggestive of a less costly process of updating of predictions about categorical information. That said, it remains to be determined if the prediction strategy differences we observe in this window are a specific result of - or at least influenced by - the audiovisual nature of our stimuli, or if these group differences would have been observed in response to audio-only or visual-only target stimuli as well.

To better understand the nature of prediction strategy differences between groups, we conducted an exploratory investigation to compare ERP responses to the first presentation of each target and the second presentation of each target. The logic behind this analysis was that if responses varied between the first and second presentation of each target, this could suggest that subjects used different information (or formed predictions at a different level) depending on whether or not they had seen the target before. For example, if a child is attending to global task-level information, they may recognize that each target is presented twice: once in a congruent (match - respond “yes”) context and once in an incongruent (cohort, rhyme, or unrelated - all respond “no”) context. Children who are attuned to this level of information during the task may form different predictions between first and second presentations of targets; on the other hand, children who are basing their predictions on trial-specific information such as the phonological and semantic content of primes and targets may not show a difference between the first and second presentation of targets. The children in this latter group may be actively forming predictions based on the context established by a trial’s prime (despite the costliness of this approach), regardless of task-level global knowledge of whether a prime has been presented before.

Waveforms for each group comparing responses time-locked to the first versus second AV target presentation are presented in **Appendix H**. Visual inspection of these waveforms did not suggest that either group had a marked difference in their neural responses between the first time they viewed an AV target versus the second time. We further conducted t-tests in the N280/P300 window of 200-350 ms within each group, comparing mean amplitude between first and second presentation. This window was used because, as discussed, this may be the timeframe in which differences in prediction strategies are manifesting in subjects’ ERP waveforms as either an N280 or P300 response. These t-tests did not suggest that either groups’ first and second presentation responses significantly differed (CI: $t(154) = -1.82$, $p = .07$; TH: $t(262) = -1.11$, $p = .27$). In turn, we can not infer that either group was utilizing different information to form their predictions between the first presentation versus the second presentation of the AV targets. It is possible that this measure was not the best way to capture changes in the nature of subjects' predictions; perhaps the first versus second half of all trials would be a better index of how subjects' strategies may change over the course of the task.

There are limitations to this study. First, we acknowledge the lack of power that this study had, in particular with our CI group, which may have limited our ability to observe the effects we initially predicted[8]. Second, there were concerns about the length of the study while balancing the number of trials per semantic-relatedness condition that could be included. Given these concerns, we presently decided not to test our subjects on trials with audio-only or visual-only manipulations, and as discussed, this limits our understanding of the unique contributions of each modality to the observed pattern of results. While we recognize this limitation, we argue that there are already many existing investigations into the role of audio-only speech cues on spoken word recognition processing, and that these findings can be used to guide our understanding of single-modality contributions from auditory input. The role of visual-only input has been studied less thoroughly, but this type of stimuli may not be a naturalistic proxy to the process of everyday spoken word recognition.

Third, as these data are cross-sectional in nature, we do not have direct measures of the development of AV speech processing at the individual level. This limits our interpretation of the proposed U-shaped trajectory of children's use of audio and visual speech cues concurrently

[8] We note, however, that Desroches et al. (2009) reported data from 15 adults; Desroches et al. (2013) reported data from 14 children with dyslexia and 15 age-matched typically-developing controls; and Malins et al. (2013) reported data from 14 children with SLI and 14 age-matched typically-developing controls. While a power analysis for the present study recommended a N of 15 for each group in order to observe the effects of interest, we suspect the number of subjects in our study was not the sole reason some of our results differed from earlier versions of this experiment.

(Jerger et al., 2009), though we recognize that the negative correlation between chronological age and early component latency is a finding that is directionally supported by this theory.

Fourth, there is some naturally-occurring variability in the timing of the onset of the audio speech information in our stimuli, due to the desire to keep the onset of visual speech information consistent (e.g., have the speaker's mouth open at the same time in each video, despite differences in when the sound began). This conscious choice strengthened the ability to detect latency differences across visually-evoked components such as the P1, but might have masked latency differences in auditory-evoked components like the N1 and later components. Indeed, the lack of strong N400 componentry observed in both groups' waveforms (especially the TH group) may be a result of inconsistent auditory word onset in our targets.[9] Moving forward, iterations of AV speech tasks such as this may opt to maintain consistency in auditory speech information onset, and this would likely create greater opportunities for alignment of both early sensory auditory components and later cognitive components.

Another limitation is that it would have been useful to collect systematic input from children during a debrief post-experiment, to potentially uncover the use of any strategies during the task. This would have been an excellent opportunity to understand if children in either group were aware of the speech sound manipulations across the conditions, or if they detected the equal likelihood of match vs mismatch trials. Without a systematic debriefing procedure in place, we were unable to collect this information in a purposeful and useful manner.

Next, we acknowledge that a recurring issue in speech processing research is task

[9] While it was beyond the scope of the present analysis to account for trial level differences in continuous variables such as audio onset or duration, this information is provided in **Appendix A** and may be the focus of a future analysis.

validity; the present pattern of results reflects neural processes that children use to match pictures with videos of a speaker naming the pictures in isolation, and to make a mouse-click response based on their judgment of whether the picture and spoken word match in meaning. It is worth noting that this picture-audiovisual word matching task that occurs in a quiet laboratory setting is a highly-controlled approximation of what would be observed in a naturalistic setting, such as a child's experience understanding their teacher in a busy classroom. There may be many other influences on how a child recognizes an audiovisual spoken word, such speaker familiarity and the context of the broader situation and discourse. We are unable to account for these influences in our task but recognize that these are factors that could particularly affect CI-using children when they are understanding spoken language.

Some directions for future research have been mentioned already in this discussion, especially with regard to disentangling the role of visual and audio speech cues during children's recognition of spoken words and to prioritizing aligning auditory speech cue onset over visual speech cue onset. It is also worthwhile to understand the role of listening effort through a cochlear implant on this process and its neural mechanisms. By applying a vocoded-speech filter to the audio stimuli used in our experiment, we can emulate the experience of hearing through a CI and have children with TH perform this version of the task. It would be worthwhile to compare the results of TH children who listen to vocoded speech stimuli with those of children who use CIs. Similarities observed between these two subject groups could draw attention to the unique effect that use of a device like a CI has on the process of audiovisual spoken word recognition. Similar approaches have been taken in other studies (see earlier introduction of Huyse et al., 2013), but more research is needed to fully understand the effects of CI-listening effort on neural mechanisms of spoken language processing.

Finally, to address how audiovisual speech recognition processes may change over time in both children with CIs and typical hearing, it would be important to implement a longitudinal version of this study. This would also allow us to gain more insight into the role of listening experience, CI-use, oral/speech therapy, and other factors as these develop over time.

In sum, the present study suggests that children who hear through cochlear implants and those who have typical hearing may not show differences in their early sensory processing of audio and visual speech cues, nor may they differ in the lexical-semantic integrative processes associated with recognizing audiovisual speech. We note, however, that group differences in intermediate-level processes associated with prediction of phonological information or task-level probabilities may manifest as different strategies employed by these two groups. These strategy differences can be reflected in the behavioral results, where both groups performed similarly in accuracy but the CI-using subjects were slower in their responses overall. The nature of these strategy differences remains to be fully explored, such that it is not yet clear if these differences occur as a result of having access to both the audio and visual speech information, or if differences occur due to task-specific demands. More work is needed to understand what factors lead to the use of different strategies during children’s audiovisual spoken word recognition, and if the targeted use or strengthening of specific strategies could be used to benefit CI-using children during this process.

References:

Abrahamse, R., Beynon, A., & Piai, V. (2021). Long-term auditory processing outcomes in early implanted young adults with cochlear implants: The mismatch negativity vs. P300 response. *Clinical Neurophysiology*, *132*(1), 258–268. https://doi.org/10.1016/j.clinph.2020.09.022

Alsius, A., Möttönen, R., Sams, M. E., Soto-Faraco, S., & Tiippana, K. (2014). Effect of attentional load on audiovisual speech perception: Evidence from ERPs. *Frontiers in Psychology*, *5*(JUL), 1–9. https://doi.org/10.3389/fpsyg.2014.00727

Andersson, A., Sanders, L. D., Coch, D., Karns, C. M., & Neville, H. J. (2018). Anterior and posterior erp rhyming effects in 3- to 5-year-old children. *Developmental Cognitive Neuroscience*, *30*(February), 178–190. https://doi.org/10.1016/j.dcn.2018.02.011

Baart, M., & Samuel, A. G. (2015). Turning a blind eye to the lexicon: ERPs show no cross-talk between lip-read and lexical context during speech sound processing. *Journal of Memory and Language*, *85*(August), 42–59. https://doi.org/10.1016/j.jml.2015.06.008

Baart, M., Stekelenburg, J. J., & Vroomen, J. (2014). Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia*, *53*(1), 115–121. https://doi.org/10.1016/j.neuropsychologia.2013.11.011

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition*, *14*(2), 201–212. https://doi.org/10.1016/0278-2626(90)90029-N

Barutchu, A., Danaher, J., Crewther, S. G., Innes-Brown, H., Shivdasani, M. N., & Paolini, A. G. (2010). Audiovisual integration in noise by children and adults. *Journal of Experimental Child Psychology*, *105*(1–2), 38–50. https://doi.org/10.1016/J.JECP.2009.08.005

Beauchamp, M. S. (2016). Audiovisual Speech Integration: Neural Substrates and Behavior. In

G. Hickock & S. L. Small (Eds.), *Neurobiology of Language* (pp. 515–526). Elsevier Inc. https://doi.org/10.1016/B978-0-12-407794-2.00042-0

Bell, N., Angwin, A. J., Arnott, W. L., & Wilson, W. J. (2019). Semantic processing in children with cochlear implants: Evidence from event-related potentials. *Journal of Clinical and Experimental Neuropsychology*, *41*(6), 576–590. https://doi.org/10.1080/13803395.2019.1592119

Bergeson, T. R., Houston, D. M., & Miyamoto, R. T. (2010). Effects of congenital hearing loss and cochlear implantation on audiovisual speech perception in infants and children. *Restorative Neurology and Neuroscience*, *28*(2), 157–165. https://doi.org/10.3233/RNN-2010-0522

Bergeson, T. R., Pisoni, D. B., & Davis, R. A. O. (2001). A longitudinal study of audiovisual speech perception by children with hearing loss who have cochlear implants. *Volta Review*, *103*(4), 347–370.

Bergeson, T. R., Pisoni, D. B., & Davis, R. A. O. (2005). Development of Audiovisual Comprehension Skills in Prelingually Deaf Children With Cochlear Implants. *Ear and Hearing*, *26*(2), 149–164. https://doi.org/10.7748/ns.26.40.25.s28

Bernstein, L. E., Auer, E. T., & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication*, *44*(1-4 SPEC. ISS.), 5–18. https://doi.org/10.1016/j.specom.2004.10.011

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, *20*(8), 2225–2234. https://doi.org/10.1111/j.1460-9568.2004.03670.x

Beynon, A. J., & Snik, A. F. M. (2004). Use of the event-related P300 potential in cochlear implant subjects for the study of strategy-dependent speech processing. *International Journal of Audiology*, *43*(SUPPL. 1).

Bonte, M. L., & Blomert, L. (2004). Developmental dyslexia: ERP correlates of anomalous phonological processing during spoken word recognition. *Cognitive Brain Research*, *21*(3), 360–376. https://doi.org/10.1016/j.cogbrainres.2004.06.010

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, *136*, 135–149. https://doi.org/10.1016/j.cognition.2014.10.017

Brothers, T., Dave, S., Hoversten, L. J., Traxler, M. J., & Swaab, T. Y. (2019). Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes. *Neuropsychologia*, *135*, 107225.

Brunellière, A., Delrue, L., & Auran, C. (2020). The contribution of audiovisual speech to lexical-semantic processing in natural spoken sentences. *Language, Cognition and Neuroscience*, *35*(6), 694–711. https://doi.org/10.1080/23273798.2019.1641612

Brydges, C. R., Fox, A. M., Reid, C. L., & Anderson, M. (2014). Predictive validity of the N2 and P3 ERP components to executive functioning in children: A latent-variable analysis. *Frontiers in Human Neuroscience*, *8*(1 FEB), 1–10. https://doi.org/10.3389/fnhum.2014.00080

Campbell, J., & Sharma, A. (2016). Visual cross-modal re-organization in children with cochlear implants. *PLoS ONE*, *11*(1), 1–18. https://doi.org/10.1371/journal.pone.0147793

Cheour, M., Leppänen, P. H., & Kraus, N. (2000). Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in infants and children. *Clinical neurophysiology*, *111*(1), 4-16.

Coch, D., Grossi, G., Coffey-Corina, S., Holcomb, P. J., & Neville, H. J. (2002). A developmental investigation of ERP auditory rhyming effects. *Developmental Science*, *5*(4), 467–489. https://doi.org/10.1111/1467-7687.00241

Coch, D., Maron, L., Wolf, M., & Holcomb, P. J. (2002). Word and picture processing in children: An event-related potential study. *Developmental Neuropsychology*, *22*(1), 373–406. https://doi.org/10.1207/S15326942dn2201_3

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*(3), 256–266. https://doi.org/10.1162/jocn.1994.6.3.256

Connor, C. M. D., Craig, H. K., Raudenbush, S. W., Heavner, K., & Zwolan, T. A. (2006). The age at which young deaf children receive cochlear implants and their vocabulary and speech-production growth: Is there an added value for early implantation? *Ear and Hearing*, *27*(6), 628–644. https://doi.org/10.1097/01.aud.0000240640.59205.42

Corina, D. P., Blau, S., LaMarr, T., Lawyer, L. A., & Coffey-Corina, S. (2017). Auditory and visual electrophysiology of deaf children with cochlear implants: Implications for cross-modal plasticity. *Frontiers in Psychology*, *8*(FEB), 1–13. https://doi.org/10.3389/fpsyg.2017.00059

Corina, D. P., Coffey-Corina, S., Pierotti, E., Bormann, B., Lamarr, T., Lawyer, L., Backer, K. C., & Miller, L. M. (2022). Electrophysiological Examination of Ambient Speech Processing in Children With Cochlear Implants. *Journal of Speech, Language, and Hearing Research*, *65*(9), 3502–3517. https://doi.org/10.1044/2022_JSLHR-22-00004

Corina, D. P., Coffey-Corina, S., Pierotti, E., Mankel, K., & Miller, L. M. (2024). Electrophysiological study of visual processing in children with cochlear implants.

Neuropsychologia, *194*, 108774. https://doi.org/10.1016/j.neuropsychologia.2023.108774

Cummings, A., Čeponiene, R., Dick, F., Saygin, A. P., & Townsend, J. (2009). A developmental ERP study of verbal and non-verbal semantic processing. *Brain Research*, *1208*, 137–149. https://doi.org/10.1016/j.brainres.2008.02.015.A

Danielson, D. K., Bruderer, A. G., Kandhadai, P., Vatikiotis-Bateson, E., & Werker, J. F. (2017). The organization and reorganization of audiovisual speech perception in the first year of life. *Cognitive Development*, *42*, 37–48. https://doi.org/10.1016/j.cogdev.2017.02.004

de Schonen, S., Bertoncini, J., Petroff, N., Couloigner, V., & Van Den Abbeele, T. (2018). Visual cortical activity before and after cochlear implantation: A follow up ERP prospective study in deaf children. *International Journal of Psychophysiology*, *123*, 88–102. https://doi.org/10.1016/j.ijpsycho.2017.10.009

Delaney-Busch, N., Morgan, E., Lau, E. F., & Kuperberg, G. R. (2019). Neural evidence for Bayesian trial-by-trial adaptation on the N400 during semantic priming. *Cognition*, *187*, 10–20. https://doi.org/10.1016/j.cognition.2019.01.001

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. In *Journal of Neuroscience Methods* (Vol. 134). http://www.sccn.ucsd.edu/eeglab/

DesJardins, J. L., & Eisenberg, L. S. (2007). Maternal contributions: Supporting language development in young children with cochlear implants. *Ear and Hearing*, *28*(4), 456–469. https://doi.org/10.1097/AUD.0b013e31806dc1ab

Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, *21*(10), 1893–1906.

https://doi.org/10.1162/jocn.2008.21142

Desroches, A. S., Newman, R. L., Robertson, E. K., & Joanisse, M. F. (2013). Electrophysiological indices of phonological impairments in dyslexia. *Journal of Speech, Language, and Hearing Research*, *56*(1), 250–264. https://doi.org/10.1044/1092-4388(2012/10-0351)

Dossey, E., Jones, Z., & Clopper, C. G. (2023). Relative Contributions of Social, Contextual, and Lexical Factors in Speech Processing. *Language and Speech*, *66*(2), 322–353. https://doi.org/10.1177/00238309221107870

Dunn, C. C., Walker, E. A., Oleson, J., Kenworthy, M., Van Voorst, T., Tomblin, J. B., Ji, H., Kirk, K. I., Mcmurray, B., Hanson, M., & Gantz, B. J. (2014). Longitudinal speech perception and language performance in pediatric cochlear implant users: The effect of age at implantation. *Ear and Hearing*, *35*(2), 148–160. https://doi.org/10.1097/AUD.0b013e3182a4a8f0

Duta, M. D., Styles, S. J., & Plunkett, K. (2012). ERP correlates of unexpected word forms in a picture-word study of infants and adults. *Developmental Cognitive Neuroscience*, *2*(2), 223–234. https://doi.org/10.1016/j.dcn.2012.01.003

Eddine, S. N., Brothers, T., Wang, L., Spratling, M., & Kuperberg, G. R. (2024). A predictive coding model of the N400. *Cognition*, *246*, 105755

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience*, *2*(10), 704–716. https://doi.org/10.1038/35094565

Eisenberg, L. S., Shannon, R. V., Schaefer Martinez, A., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *Journal of Acoustical*

Society of America, *107*, 2704–2710. https://doi.org/10.1037/h0073226

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469-495.

Fisher, C. G. (1968). Confusions among visually perceived consonants. *Journal of Speech and Hearing Research*, *11*(4), 796–804. https://doi.org/10.1044/jshr.1104.796

Folstein, J. R., & Van Petten, C. (2011). After the P3: late executive processes in stimulus categorization. *Psychophysiology*, *48*(6), 825-841.

Fort, M., Kandel, S., Chipot, J., Savariaux, C., Granjon, L., & Spinelli, E. (2013). Seeing the initial articulatory gestures of a word triggers lexical access. *Language and Cognitive Processes*, *28*(8), 1207–1223. https://doi.org/10.1080/01690965.2012.701758

Frauenfelder, U. H., & Tyler, L. K. (1987). The process of spoken word recognition: An introduction. *Cognition*, *25*(1–2), 1–20. https://doi.org/10.1016/0010-0277(87)90002-3

Friedrich, M., & Friederici, A. D. (2004). N400-like Semantic Incongruity Effect in 19-Month-Olds: Processing Known Words in Picture Contexts. *Journal of Cognitive Neuroscience*, *16*(8), 1465–1477. https://www.mitpressjournals.org/doi/pdfplus/10.1162/0898929042304705

Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology*, *22*(7), 615–621. https://doi.org/10.1016/j.cub.2012.02.015

Geers, A. E., Brenner, C. A., & Davidson, L. S. (2003). Factors associated with development of speech perception skills in children implanted by age five. *Ear and Hearing*, *24*(1 SUPPL.). https://doi.org/10.1097/01.aud.0000051687.99218.0f

Geers, A. E., & Hayes, H. (2011). Reading, writing, and phonological processing skills of adolescents with 10 or more years of cochlear implant experience. *Ear and Hearing*, *32*(1 Suppl), 0–23. https://doi.org/10.1097/aud.0b013e3181fa41fa

Geers, A. E., Strube, M., Tobey, E. A., & Pisoni, D. B. (2011). Epilogue: Factors Contributing to Long-Term Outcomes of Cochlear Implantation in Early Childhood. *Ear and Hearing*, *32*, 84–92. https://doi.org/10.1097/AUD.0b013e3181ffd5b5

Hall, W. C. (2017). What You Don't Know Can Hurt You: The Risk of Language Deprivation by Impairing Sign Language Development in Deaf Children. *Maternal and Child Health Journal*, *21*(5), 961–965. https://doi.org/10.1007/s10995-017-2287-y

Hauthal, N., Thorne, J. D., Debener, S., & Sandmann, P. (2014). Source localisation of visual evoked potentials in congenitally deaf individuals. *Brain Topography*, *27*(3), 412–424. https://doi.org/10.1007/s10548-013-0341-7

Henderson, L. M., Baseler, H. A., Clarke, P. J., Watson, S., & Snowling, M. J. (2011). The N400 effect in children: Relationships with comprehension, vocabulary and decoding. *Brain and Language*, *117*(2), 88–99. https://doi.org/10.1016/j.bandl.2010.12.003

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech understanding. *Nature*, *8*(May), 393–402. www.nature.com/reviews/neuro%0Ahttps://www-nature-com.ezp-prod1.hul.harvard.edu/articles/nrn2113.pdf

Hickok, G. (2012). The cortical organization of speech processing: feedback control and predictive coding the context of a dual-stream model. Journal of Communication Disorders, 45(6), 3. *Journal of Communication Disorders*, *45*(6), 393–402. https://doi.org/10.1016/j.jcomdis.2012.06.004

Hickok, G. (2022). The dual stream model of speech and language processing. In *Handbook of*

Clinical Neurology (1st ed., Vol. 185, Issue C). Elsevier B.V. https://doi.org/10.1016/B978-0-12-823384-9.00003-7

Huyse, A., Berthommier, F., & Leybaert, J. (2013). Degradation of labial information modifies audiovisual speech perception in cochlear-implanted children. *Ear and Hearing*, *34*(1), 110–121. https://doi.org/10.1097/AUD.0b013e3182670993

Jerger, S., Damian, M. F., Spence, M. J., Tye-Murray, N., & Abdi, H. (2009). Developmental shifts in children's sensitivity to visual speech: A new multimodal picture word task. *Journal of Experimental Child Psychology*, *102*(1), 40–59. https://doi.org/10.1016/j.jecp.2008.08.002.Developmental

Joanisse, M. F., & McClelland, J. L. (2015). Connectionist perspectives on language learning, representation and processing. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(3), 235–247. https://doi.org/10.1002/wcs.1340

Johnson, R. (1988). The amplitude of the P300 component of the event-related potential: Review and synthesis. *Advances in psychophysiology*, *3*(April), 69-137.

Kaganovich, N., Schumaker, J., Macias, D., & Gustafson, D. (2015). Processing of audiovisually congruent and incongruent speech in school-age children with a history of specific language impairment: A behavioral and event-related potentials study. *Developmental Science*, *18*(5), 751–770. https://doi.org/10.1111/desc.12263

Kaganovich, N., Schumaker, J., & Rowland, C. (2016). Matching heard and seen speech: An ERP study of audiovisual word recognition. *Brain and Language*, *157–158*, 14–24. https://doi.org/10.1016/j.bandl.2016.04.010

Kaganovich, N., Schumaker, J., & Rowland, C. (2016). Atypical audiovisual word processing in school-age children with a history of specific language impairment: An event-related potential study. *Journal of Neurodevelopmental Disorders*, *8*(1), 1–22. https://doi.org/10.1186/s11689-016-9168-3

Kallioinen, P., Olofsson, J., Nakeva Von Mentzer, C., Lindgren, M., Ors, M., Sahlén, B. S., Lyxell, B., Engström, E., & Uhlén, I. (2016). Semantic processing in deaf and hard-of-hearing children: Large N400 mismatch effects in brain responses, despite poor semantic ability. *Frontiers in Psychology*, *7*(AUG), 1–10. https://doi.org/10.3389/fpsyg.2016.01146

Knowland, V. C. P., Mercure, E., Karmiloff-Smith, A., Dick, F., & Thomas, M. S. C. (2014). Audio-visual speech perception: A developmental ERP investigation. *Developmental Science*, *17*(1), 110–124. https://doi.org/10.1111/desc.12098

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*(4577), 1138–1141. https://doi.org/10.1126/science.7146899

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M., & Federmeier, K. D. (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. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, *207*(4427), 203–205.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science*, *197*(4305), 792-795.

Lachs, L., Pisoni, D. B., & Kirk, K. I. (2001). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. *Ear and Hearing*, *22*(3), 236–251. https://doi.org/10.1097/00003446-200106000-00007

Lalonde, K., & Holt, R. (2015). Preschoolers benefit from visually salient speech cues. *Journal of Speech, Language, and Hearing Research*, *58*, 135–15. https://doi.org/10.1044/2014

Lalonde, K., & McCreery, R. W. (2020). Audiovisual Enhancement of Speech Perception in Noise by School-Age Children Who Are Hard of Hearing. *Ear and Hearing*, 705–719. https://doi.org/10.1097/AUD.0000000000000830

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews. Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Lederberg, A. R., Schick, B., & Spencer, P. E. (2013). Language and literacy development of deaf and hard-of-hearing children: successes and challenges. In *Developmental psychology* (Vol. 49, Issue 1, pp. 15–30). https://doi.org/10.1037/a0029558

Leibold, L. J., & Buss, E. (2019). Masked Speech Recognition in School-Age Children. *Frontiers in Psychology*, *10*(September). https://doi.org/10.3389/fpsyg.2019.01981

Levi-Aharoni, H., Shriki, O., & Tishby, N. (2020). Surprise response as a probe for compressed memory states. *PLoS Computational Biology*, *16*(2), e1007065. https://doi.org/10.1371/journal.pcbi.1007065

Liu, J., Liang, M., Chen, Y., Wang, Y., Cai, Y., Chen, S., Chen, L., Li, X., Qiu, Z., Jiang, J., Wang, J., & Zheng, Y. (2017). Visual cortex activation decrement following cochlear implantation in prelingual deafened children. *International Journal of Pediatric Otorhinolaryngology*, *99*(2017), 85–89. https://doi.org/10.1016/j.ijporl.2017.04.011

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213.

https://doi.org/10.3389/fnhum.2014.00213

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36. https://doi.org/10.1097/00003446-199802000-00001

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, *4*(11), 432–440. https://doi.org/10.1016/S1364-6613(00)01545-X

Maess, B., Mamashli, F., Obleser, J., Helle, L., & Friederici, A. D. (2016). Prediction signatures in the brain: Semantic pre-activation during language comprehension. *Frontiers in Human Neuroscience*, *10*(NOV2016). https://doi.org/10.3389/fnhum.2016.00591

Malins, J. G., Desroches, A. S., Robertson, E. K., Newman, R. L., Archibald, L. M. D., & Joanisse, M. F. (2013). ERPs reveal the temporal dynamics of auditory word recognition in specific language impairment. *Developmental Cognitive Neuroscience*, *5*, 134–148. https://doi.org/10.1016/j.dcn.2013.02.005

Mangun, G. R., & Hillyard, S. A. (1990). Allocation of visual attention to spatial locations: tradeoff functions for event-related brain potentials and detection performance. *Perception & Psychophysics*, *47*(6), 532-550.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1–71. https://doi.org/10.1016/0010-0277(80)90015-3

Mattys, S. L., Bernstein, L. E., & Auer, E. T. (2002). Stimulus-based lexical distinctiveness as a general word-recognition mechanism. *Perception and Psychophysics*, *64*(4), 667–679. https://doi.org/10.3758/BF03194734

Mayberry, R. I. (2007). When timing is everything: Age of first-language acquisition effects on second-language learning. *Applied Psycholinguistics*, *28*, 537–549.

https://doi.org/10.1017.S0142716407070294

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McCreery, R. W., & Stelmachowicz, P. G. (2011). Audibility-based predictions of speech recognition for children and adults with normal hearing. *The Journal of the Acoustical Society of America*, *130*(6), 4070-4081.

McCreery, R. W., Spratford, M., Kirby, B., & Brennan, M. (2017). Individual differences in language and working memory affect children's speech recognition in noise. *International Journal of Audiology*, *56*(5), 306–315. https://doi.org/10.1080/14992027.2016.1266703

McDonald, J. J., & Green, J. J. (2008). Isolating event-related potential components associated with voluntary control of visuo-spatial attention. *Brain Research*, *1227*, 96–109. https://doi.org/10.1016/j.brainres.2008.06.034

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

McMurray, B., Farris-Trimble, A., & Rigler, H. (2017). Waiting for lexical access: Cochlear implants or severely degraded input lead listeners to process speech less incrementally. *Cognition*, *169*(May 2016), 147–164. https://doi.org/10.1016/j.cognition.2017.08.013

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). Article The syllable ' s role in speech segmentation The Syllable ' s Role in Speech Segmentation. *Journal of Verbal Learning and Verbal Behavior*, *20*(3), 298–305.

Mercure, E., Kushnerenko, E., Goldberg, L., Bowden-Howl, H., Coulson, K., Johnson, M. H., & MacSweeney, M. (2019). Language experience influences audiovisual speech integration in unimodal and bimodal bilingual infants. *Developmental Science*, *22*(1), 1–9.

https://doi.org/10.1111/desc.12701

Meristo, M., Strid, K., & Hjelmquist, E. (2016). Early conversational environment enables spontaneous belief attribution in deaf children. *Cognition*, *157*, 139–145. https://doi.org/10.1016/j.cognition.2016.08.023

Moore, D. R., & Shannon, R. V. (2009). Beyond cochlear implants: Awakening the deafened brain. *Nature Neuroscience*, *12*(6), 686–691. https://doi.org/10.1038/nn.2326

Näätänen, R., & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology*, *24*(4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Nabé, M., Schwartz, J. L., & Diard, J. (2021). COSMO-Onset: A Neurally-Inspired Computational Model of Spoken Word Recognition, Combining Top-Down Prediction and Bottom-Up Detection of Syllabic Onsets. *Frontiers in Systems Neuroscience*, *15*(August), 1–21. https://doi.org/10.3389/fnsys.2021.653975

Navarra, J., & Soto-Faraco, S. (2005). Hearing lips in a second language: visual articulatory information enables the perception of second language sounds. *Psychological Research 2005 71:1*, *71*(1), 4–12. https://doi.org/10.1007/S00426-005-0031-5

Ness, T., & Meltzer-Asscher, A. (2021). Rational Adaptation in Lexical Prediction: The Influence of Prediction Strength. *Frontiers in Psychology*, *12*(April). https://doi.org/10.3389/fpsyg.2021.622873

Newman, R. L., Connolly, J. F., Service, E., & McIvor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology*, *40*(4), 640–647. https://doi.org/10.1111/1469-8986.00065

Nieuwland, M. S. (2019). Neuroscience and Biobehavioral Reviews Do ‘ early ’ brain responses reveal word form prediction during language comprehension ? A critical review. *Neuroscience and Biobehavioral Reviews*, *96*(November 2018), 367–400. https://doi.org/10.1016/j.neubiorev.2018.11.019

Niparko, J. K., Tobey, E. A., Thal, D. J., Eisenberg, L. S., Wang, N. Y., Quittner, A. L., & Fink, N. E. (2010). Spoken language development in children following cochlear implantation. *Jama*, *303*(15), 1498–1506. https://doi.org/10.1001/jama.2010.451

Nittrouer, S., Manning, C., & Meyer, G. (1993). The perceptual weighting of acoustic cues changes with linguistic experience. *The Journal of the Acoustical Society of America*, *94*(3_Supplement), 1865-1865.

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *The Journal of the Acoustical Society of America*, *112*(2), 711-719.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189–234. https://doi.org/10.1016/0010-0277(94)90043-4

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian Model of Continuous Speech Recognition. *Psychological Review*, *115*(2), 357–395. https://doi.org/10.1037/0033-295X.115.2.357

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(3), 299–370. https://doi.org/10.1017/S0140525X00003241

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory-of-mind development for children with deafness or autism. *Child Development*, *76*(2), 502–517. https://doi.org/10.1111/j.1467-8624.2005.00859.x

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Pierotti, E., Coffey-Corina, S., Schaefer, T., & Corina, D. P. (2021). Semantic word integration in children with cochlear implants: electrophysiological evidence. *Language, Cognition and Neuroscience*, *36*(10), 1–18. https://doi.org/10.1080/23273798.2021.1957954

Pierotti, E., Coffey-Corina, S., & Corina, D. P. (*in preparation*). Seeing and hearing speech: an EEG Study of audiovisual word perception in typically hearing and cochlear implant-using children.

Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, *25*(1–2), 21–52. https://doi.org/10.1016/0010-0277(87)90003-5

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, *21*(6), 322–334. https://doi.org/10.1038/s41583-020-0304-4

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Praamstra, P., & Stegeman, D. F. (1993). Phonological effects on the auditory N400 event-related brain potential. *Cognitive Brain Research*, *1*(2), 73–86. https://doi.org/10.1016/0926-6410(93)90013-U

Rabovsky, M., & McRae, K. (2014). Simulating the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition*, *132*(1), 68–89. https://doi.org/10.1016/j.cognition.2014.03.010

Rauss, K., & Pourtois, G. (2013). What is bottom-up and what is top-down in predictive coding. *Frontiers in Psychology*, *4*(MAY), 1–8. https://doi.org/10.3389/fpsyg.2013.00276

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In R. Campbell & B. Dodd (Eds.), *Hearing by Eye: The Psychology of Lip-reading* (pp. 97–113). Lawrence Erlbaum Associates, Inc.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17*(5), 1147–1153. https://doi.org/10.1093/cercor/bhl024

Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J. J. (2011). The development of multisensory speech perception continues into the late childhood years. *European Journal of Neuroscience*, *33*(12), 2329–2337. https://doi.org/10.1111/j.1460-9568.2011.07685.x

Sandmann, P., Dillier, N., Eichele, T., Meyer, M., Kegel, A., Pascual-Marqui, R. D., Marcar, V. L., Jäncke, L., & Debener, S. (2012). Visual activation of auditory cortex reflects maladaptive plasticity in cochlear implant users. *Brain*, *135*(2), 555–568. https://doi.org/10.1093/brain/awr329

Schorr, E. A., Fox, N. A., Van Wassenhove, V., & Knudsen, E. I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(51), 18748–18750. https://doi.org/10.1073/pnas.0508862102

Schorr, E. A., Roth, F. P., & Fox, N. A. (2008). A Comparison of the Speech and Language Skills of Children With Cochlear Implants and Children With Normal Hearing. *Communication Disorders Quarterly*, *29*(4), 195–210. https://doi.org/10.15724/jslhd.2020.29.1.013

Schneider, J. M., Poudel, S., Abel, A. D., & Maguire, M. J. (2023). Age and vocabulary knowledge differentially influence the N400 and theta responses during semantic retrieval.

Developmental Cognitive Neuroscience, *61*, 101251.

Schwartz, J. L., & Savariaux, C. (2014). No, There Is No 150 ms Lead of Visual Speech on Auditory Speech, but a Range of Audiovisual Asynchronies Varying from Small Audio Lead to Large Audio Lag. *PLoS Computational Biology*, *10*(7). https://doi.org/10.1371/journal.pcbi.1003743

Scott, S. K. (2019). From speech and talkers to the social world: The neural processing of human spoken language. *Science*, *366*(6461), 58–62. https://doi.org/10.1126/science.aax0288

Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Developmental Science*, *11*(2), 306–320. https://doi.org/10.1111/j.1467-7687.2008.00677.x

Sharma, A., Dorman, M. F., & Spahr, A. J. (2002). A sensitive period for the development of the central auditory system in children with cochlear implants: implications for age of implantation. *Ear and hearing*, *23*(6), 532-539.

Sharma, A., Dorman, M. F., & Kral, A. (2005). The influence of a sensitive period on central auditory development in children with unilateral and bilateral cochlear implants. *Hearing Research*, *203*, 134–143. https://doi.org/10.1016/j.heares.2004.12.010

Stevenson, R. A., Sheffield, S. W., Butera, I. M., Gifford, R. H., & Wallace, M. T. (2017). Multisensory Integration in Cochlear Implant Recipients. In *Ear and hearing* (Vol. 38, Issue 5). https://doi.org/10.1097/AUD.0000000000000435

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2012). Language-related ERP components. In *Oxford handbook of event-related potential components* (pp. 397–440).

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108*(3), 850–855.

https://doi.org/10.1016/j.cognition.2008.05.009

Tye-Murray, N., Sommers, M., & Spehar, B. (2007). Auditory and Visual Lexical Neighborhoods in Audiovisual Speech Perception. *Trends in Amplification*, *11*(4), 233–241. https://doi.org/10.1177/1084713807307409

Vallat, R. (2018). Pingouin: statistics in Python. *Journal of Open Source Software*, *3*(31), 1026. https://doi.org/10.21105/joss.01026

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time Course of Word Identification and Semantic Integration in Spoken Language. *Journal of Experimental Psychology: Learning Memory and Cognition*, *25*(2), 394–417. https://doi.org/10.1037/0278-7393.25.2.394

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1181–1186. https://doi.org/10.1073/pnas.0408949102

Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(3), 387–401. https://doi.org/10.1002/wcs.1178

Wie, O. B., Falkenberg, E. S., Tvete, O., & Tomblin, B. (2007). Children with a cochlear implant: Characteristics and determinants of speech recognition, speech-recognition growth rate, and speech production. *International Journal of Audiology*, *46*(5), 232–243. https://doi.org/10.1080/14992020601182891

Yeung, H. H., & Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological science*, *24*(5), 603-612. https://doi.org/10.1177/0956797612458802

Yoshinaga-Itano, C., Baca, R. L., & Sedey, A. L. (2010). Describing the trajectory of language development in the presence of severe-to-profound hearing loss: A closer look at children with cochlear implants versus hearing aids. *Otology and Neurotology*, *31*(8), 1268–1274. https://doi.org/10.1097/MAO.0b013e3181f1ce07

Appendix A. Details of each audiovisual word target used in the experiment. Video duration is how long the target lasted in total, while the audio onset details how many ms after the onset of the video passed until meaningful audio associated with the target word began. The audio duration is how long the spoken word itself lasted. SUBTLWF is the SUBTLEX frequency per million words, and Lg10WF is the same measure of word frequency but on a logarithmic scale. Viseme is based on the visual facial shape created by the initial phoneme of the word. Viseme distinctiveness was calculated by determining how many other phonemes shared the same visual facial shape (e.g. a viseme that has two corresponding phonemes has a distinctiveness value of ½ or 0.5).

Concept	Video duration (ms)	Audio onset (ms)	Audio duration (ms)	SUBTLWF	Lg10WF	Initial phoneme	Viseme distinctiveness
BALL	1668	243	921	104.96	3.7287	/b/	0.33333333
BAT	1368	235	668	20.63	3.0224	/b/	0.33333333
BATH	1468	239	598	31.12	3.2009	/b/	0.33333333
BEACH	1568	238	821	56.63	3.4607	/b/	0.33333333
BELL	1802	245	934	39.33	3.3025	/b/	0.33333333
BLOCK	1568	249	755	40.53	3.3156	/b/	0.33333333
BOAT	1468	241	651	95.78	3.689	/b/	0.33333333
BOMB	1502	199	725	53.65	3.4373	/b/	0.33333333
BONE	1502	237	790	26.06	3.1239	/b/	0.33333333
BOOT	1602	256	819	11.14	2.7551	/b/	0.33333333
BOWL	1602	231	684	21.45	3.0394	/b/	0.33333333
BOX	1468	257	796	89.75	3.6607	/b/	0.33333333
BROW	1568	232	809	1.84	1.9777	/b/	0.33333333
BUG	1468	200	603	20.94	3.029	/b/	0.33333333
BUN	1502	229	693	2.88	2.1703	/b/	0.33333333
CAGE	1735	238	1020	20.27	3.0149	/c/	0.5
CAKE	1201	240	527	45.06	3.3615	/c/	0.5
CANE	1568	231	920	8.33	2.6513	/c/	0.5
CAP	1368	243	652	18.75	2.9809	/c/	0.5
CAPE	1301	262	544	8.24	2.6243	/c/	0.5

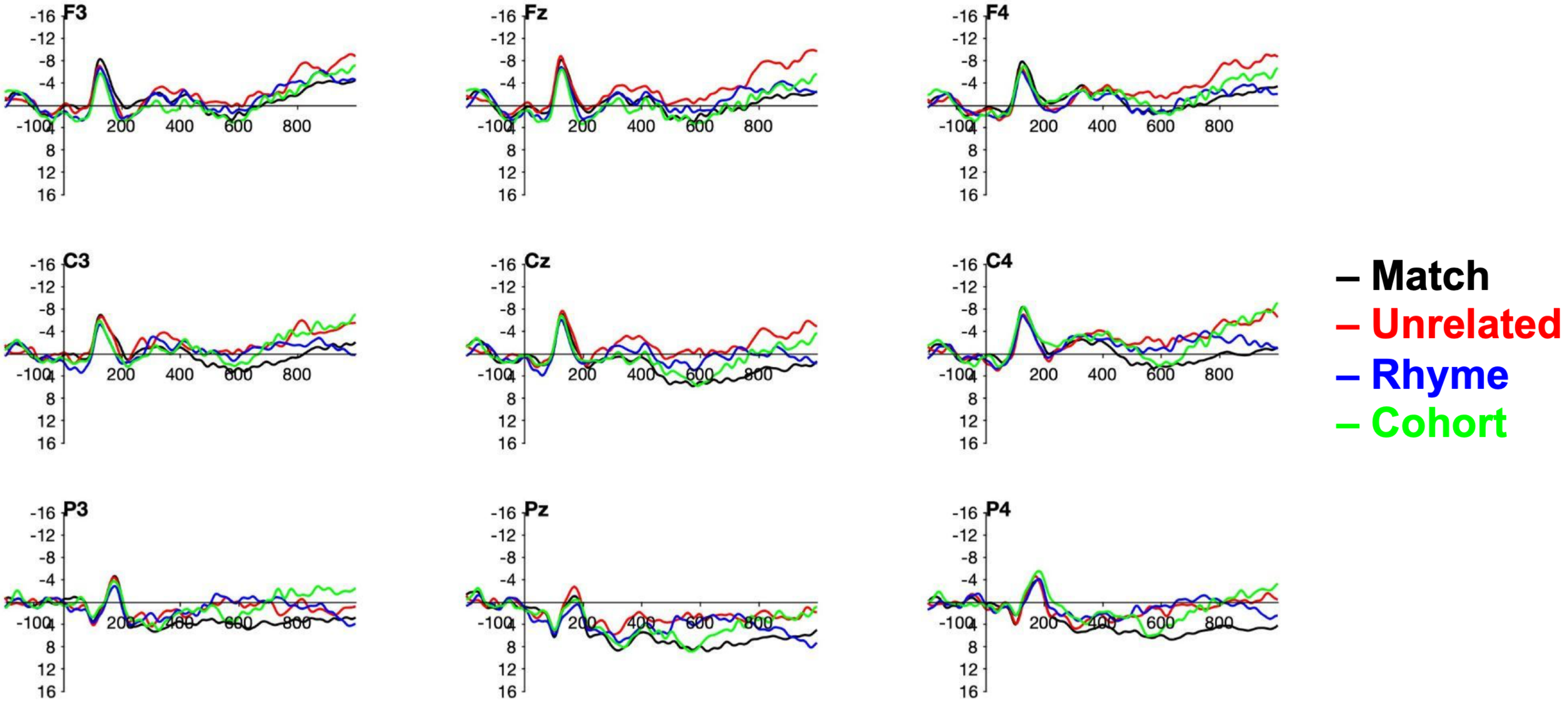
Concept	Video duration (ms)	Audio onset (ms)	Audio duration (ms)	SUBTLWF	Lg10WF	Initial phoneme	Viseme distinctiveness
CARD	1401	186	654	85.43	3.6393	/c/	0.5
CART	1502	236	727	9.04	2.6646	/c/	0.5
CAT	1535	251	712	66.33	3.5294	/c/	0.5
CHEF	1702	251	690	1.02	1.7243	/ É/	0.25
CHICK	1568	239	572	26.16	3.1255	/ ß/	0.25
CHIP	1401	241	409	20.61	3.022	/ ß/	0.25
CLOCK	1368	199	562	58.63	3.4758	/c/	0.5
CLOTH	1568	258	548	6.1	2.4942	/c/	0.5
COAT	1568	237	813	42.08	3.3318	/c/	0.5
COMB	1735	232	926	6.06	2.4914	/c/	0.5
CONE	1401	241	742	2.92	2.1761	/c/	0.5
COUCH	1668	240	890	23.47	3.0785	/c/	0.5
COW	1435	239	561	25.51	3.1146	/c/	0.5
DICE	1568	245	889	10.45	2.7275	/d/	0.5
DOG	1568	245	916	192.84	3.9928	/d/	0.5
DOLL	1635	245	769	24.76	3.1017	/d/	0.5
EYES	1635	245	959	221.55	4.0531	/a/	1
FAN	1869	233	1077	35.14	3.2536	/f/	0.5
FOX	1869	233	846	21.61	3.0426	/f/	0.5
GHOST	1735	233	850	36.59	3.2711	/g/	0.5
GOAT	1735	233	765	10.53	2.7308	/g/	0.5
HAND	1635	180	848	279.65	4.1542	/h/	1
HAT	1635	180	608	64.18	3.5151	/h/	1
HEART	1535	251	819	244.18	4.0953	/h/	1
HOSE	1535	251	1137	8.06	2.6149	/h/	1
HOUSE	1602	234	797	514	4.4185	/h/	1

Concept	Video duration (ms)	Audio onset (ms)	Audio duration (ms)	SUBTLWF	Lg10WF	Initial phoneme	Viseme distinctiveness
ICE	1602	234	751	79.55	3.6083	/a/	1
JET	1602	201	833	14.14	2.8585	§	0.25
KING	1602	201	790	129.25	3.8191	/k/	0.5
KITE	1635	236	870	2.29	2.0719	/k/	0.5
KNIFE	1635	236	740	46.8	3.378	/n/	1
KNIGHT	1468	248	794	26.76	3.1355	/n/	1
LOCK	1468	248	569	56.57	3.4603	/l/	1
LOG	1635	260	834	11.96	2.786	/l/	1
MAP	1635	260	581	31.82	3.2106	/m/	0.33333333
MASK	1902	258	1062	19.8	3.0048	/m/	0.33333333
MAT	1902	258	969	3.49	2.2529	/m/	0.33333333
MOUSE	1401	198	819	19.12	2.9894	/m/	0.33333333
MOUTH	1401	198	658	104.41	3.7264	/m/	0.33333333
MUG	1668	239	935	6.84	2.5441	/m/	0.33333333
NECK	1668	239	741	59.51	3.4823	/n/	1
NET	1468	248	725	15.55	2.8998	/n/	1
NOSE	1468	248	1179	69.75	3.5512	/n/	1
NOTE	1535	253	838	53.55	3.4365	/n/	1
NURSE	1535	253	1113	44.98	3.3608	/n/	1
PAN	1401	233	713	12.29	2.798	/p/	0.33333333
PANTS	1401	233	905	58.75	3.4767	/p/	0.33333333
PEACH	1468	226	725	6.35	2.5119	/p/	0.33333333
PEARL	1468	226	945	15.67	2.9031	/p/	0.33333333
PEAS	1635	257	963	4.65	2.3766	/p/	0.33333333
PLANE	1635	257	682	95.53	3.6878	/p/	0.33333333
PLATE	1401	243	743	25.65	3.1169	/p/	0.33333333

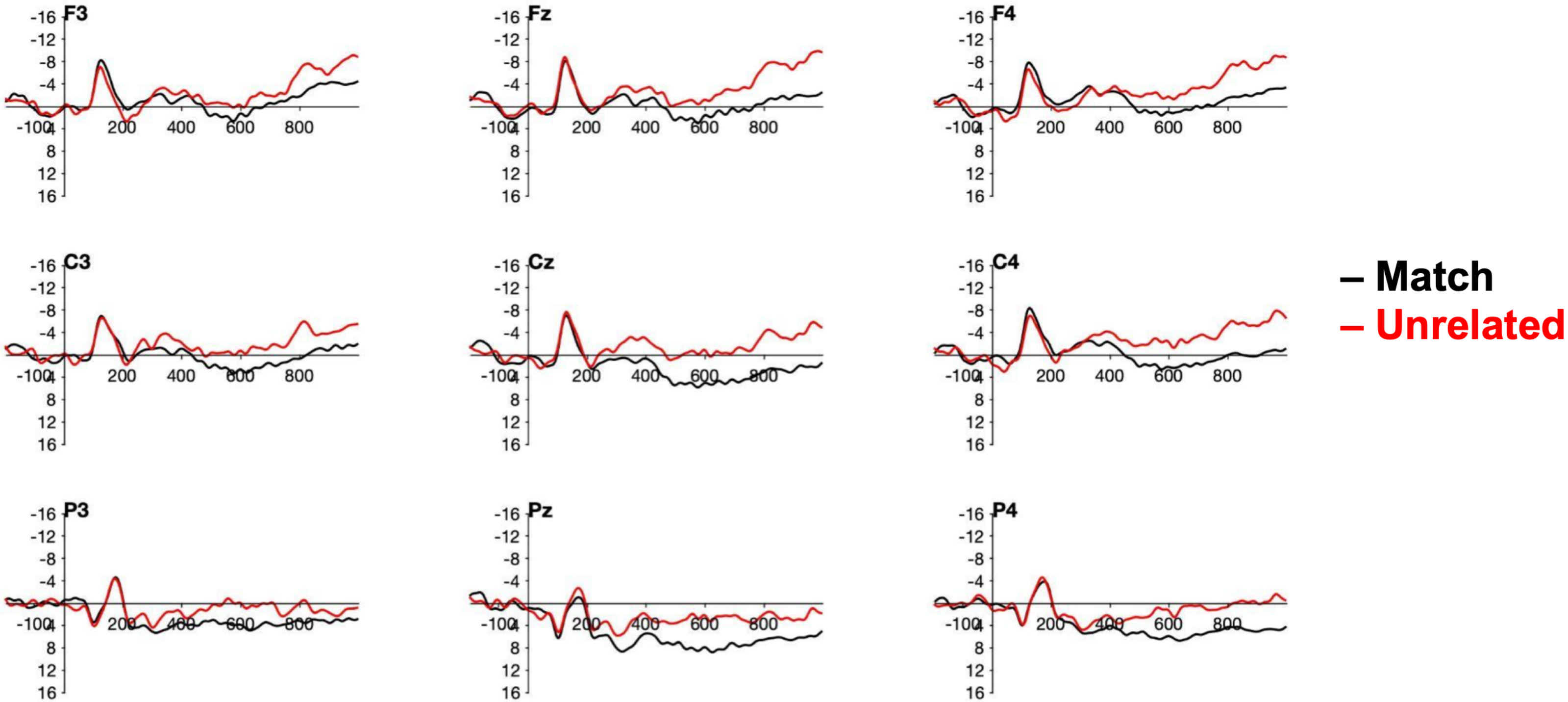
Concept	Video duration (ms)	Audio onset (ms)	Audio duration (ms)	SUBTLWF	Lg10WF	Initial phoneme	Viseme distinctiveness
PURSE	1502	284	776	19.76	3.0039	/p/	0.33333333
RAKE	1335	236	730	2.98	2.1847	/r/	1
ROAD	1335	244	787	111.94	3.7566	/r/	1
ROBE	1368	254	715	8.49	2.6375	/r/	1
ROPE	1568	242	495	22.71	3.0641	/r/	1
ROSE	1235	236	880	53.02	3.4322	/r/	1
SEAL	1635	256	623	14.75	2.8768	/s/	0.5
SEED	1401	238	897	7.57	2.5877	/s/	0.5
SHELL	1568	229	746	13.22	2.8293	/ É/	0.25
SHIP	1401	244	610	98.88	3.7028	/ É/	0.25
SOAP	1668	241	864	15.2	2.8899	/s/	0.5
SOCK	1568	227	806	8.98	2.6618	/s/	0.5
SOUP	1435	253	807	25.2	3.1092	/s/	0.5
SUIT	1602	253	754	68.61	3.5441	/s/	0.5
TAPE	1768	238	790	68.84	3.5456	/t/	0.5
TOAST	1568	231	819	33.47	3.2325	/t/	0.5
TOES	1568	231	916	12.51	2.8055	/t/	0.5
TRAIN	1368	226	993	95.06	3.6857	/t/	0.5
WHEEL	1468	234	823	27.06	3.1402	/u/	1
WIG	1602	255	781	8.35	2.6304	/u/	1
WING	1401	253	650	20.24	3.0141	/u/	1

Appendix B. Auditory-evoked waveforms by Group and Condition: All sites

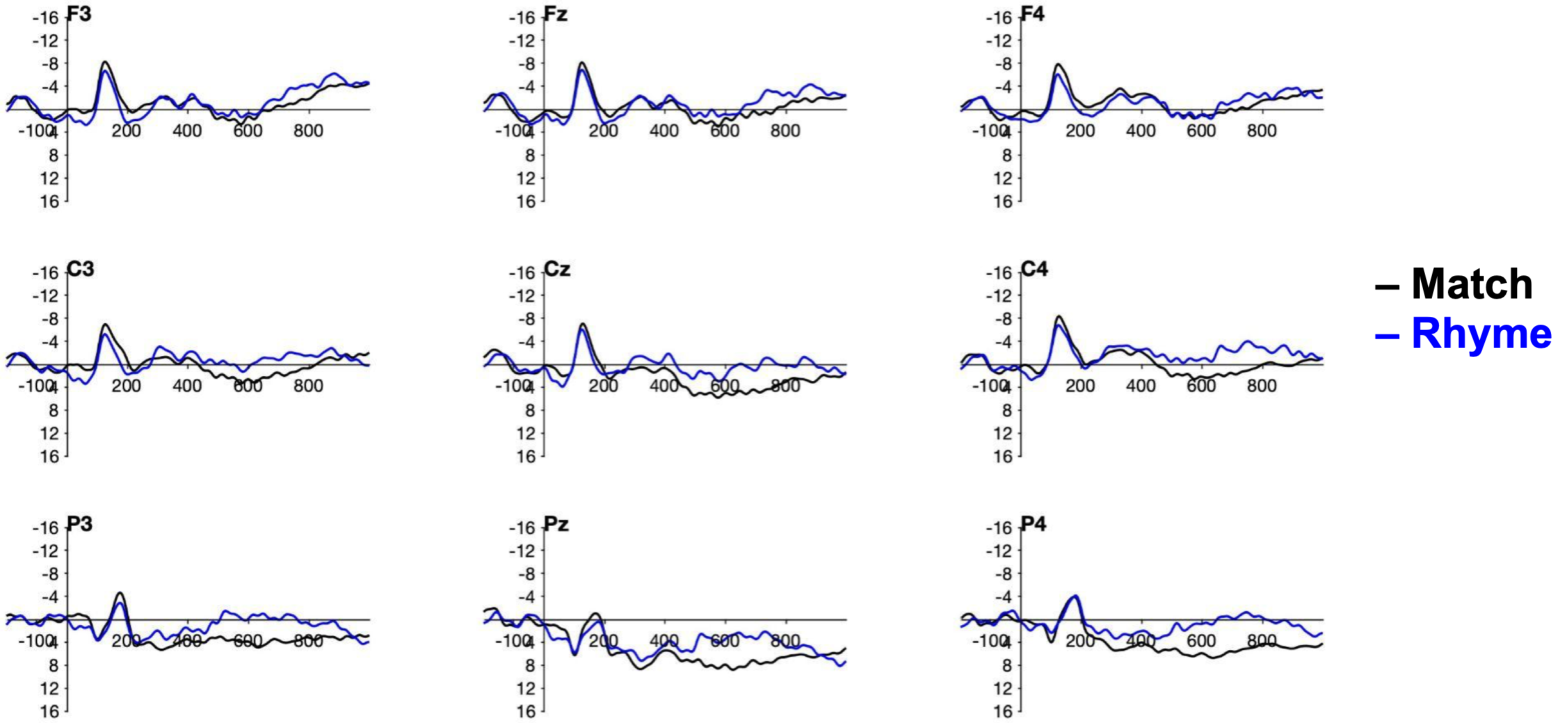
CI-Using Children (n = 13): All Conditions



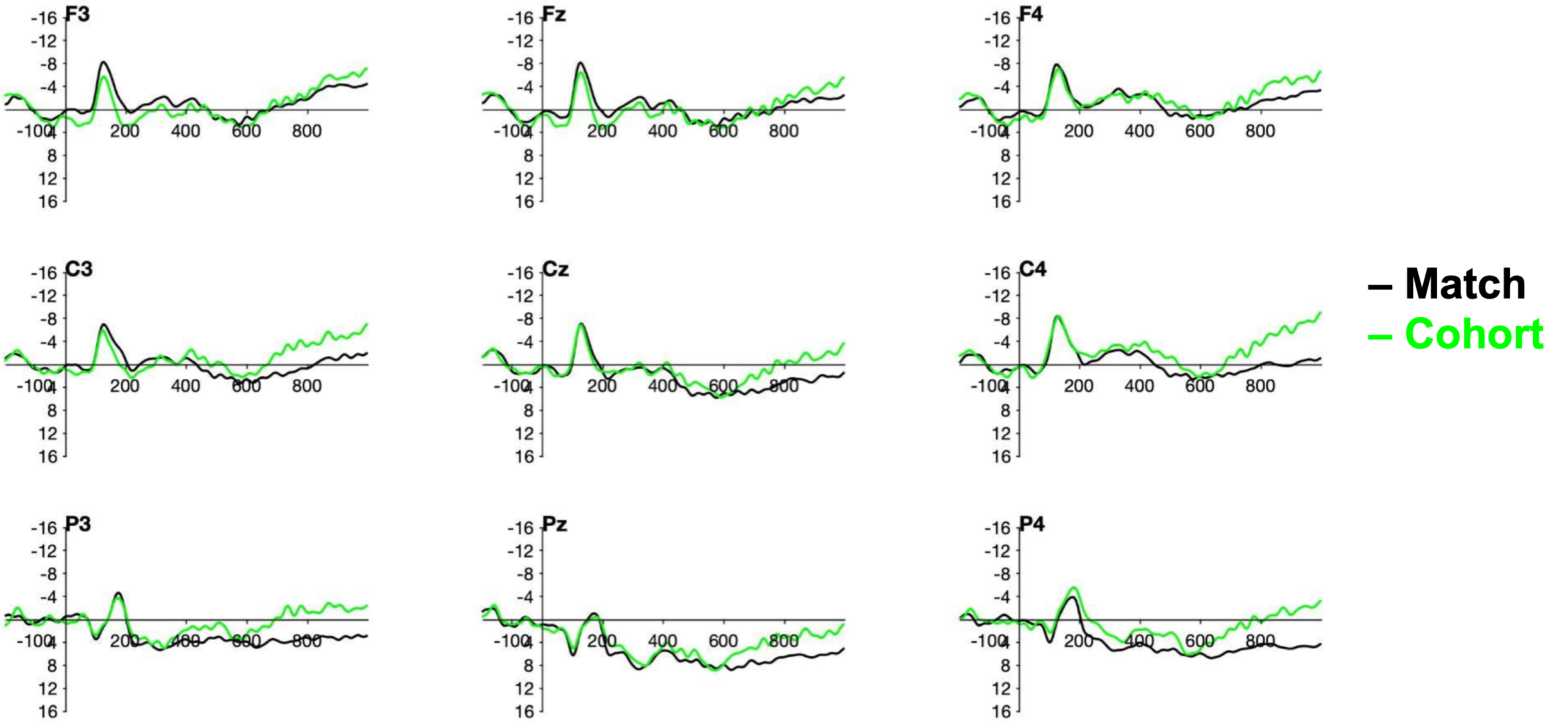
CI-Using Children (n = 13): Match Vs. Unrelated



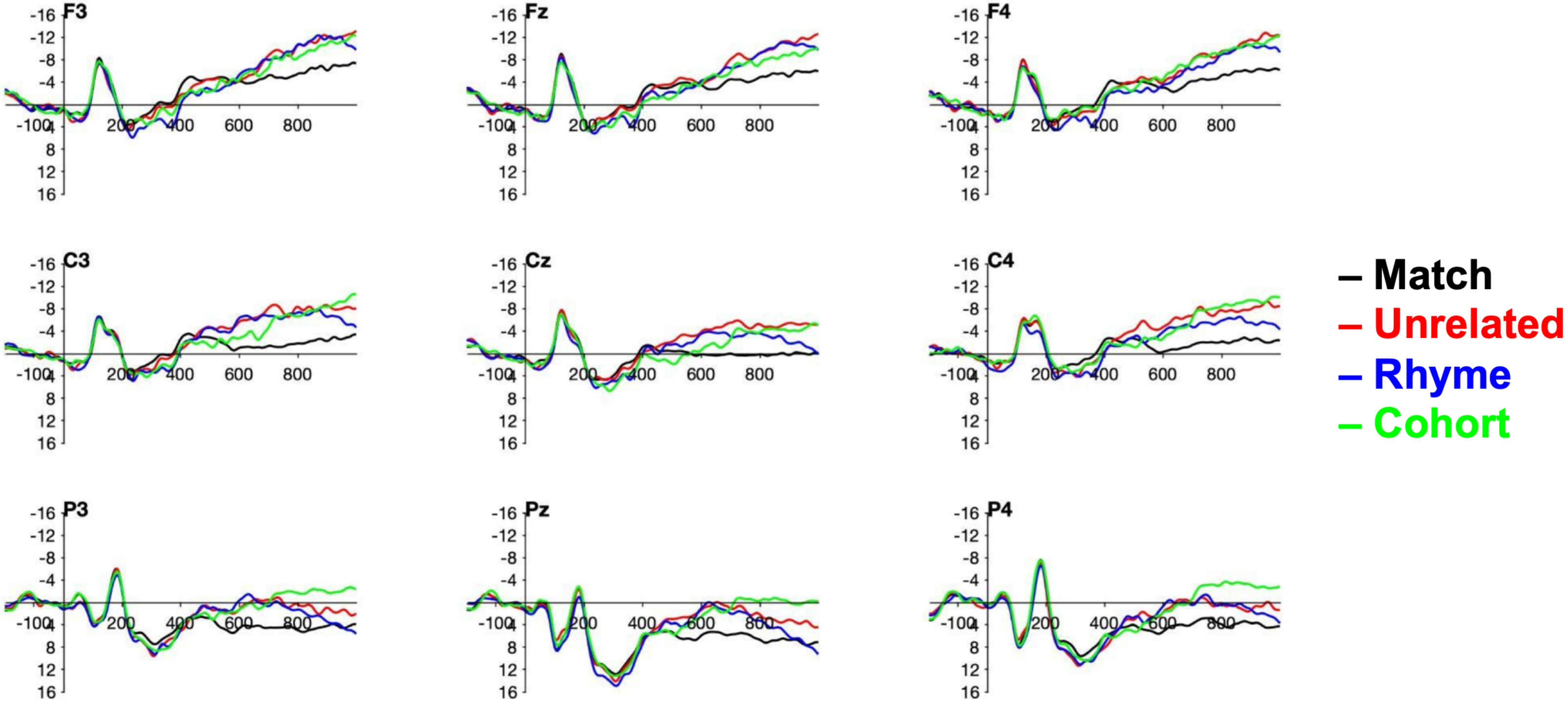
CI-Using Children (n = 13): Match Vs. Rhyme



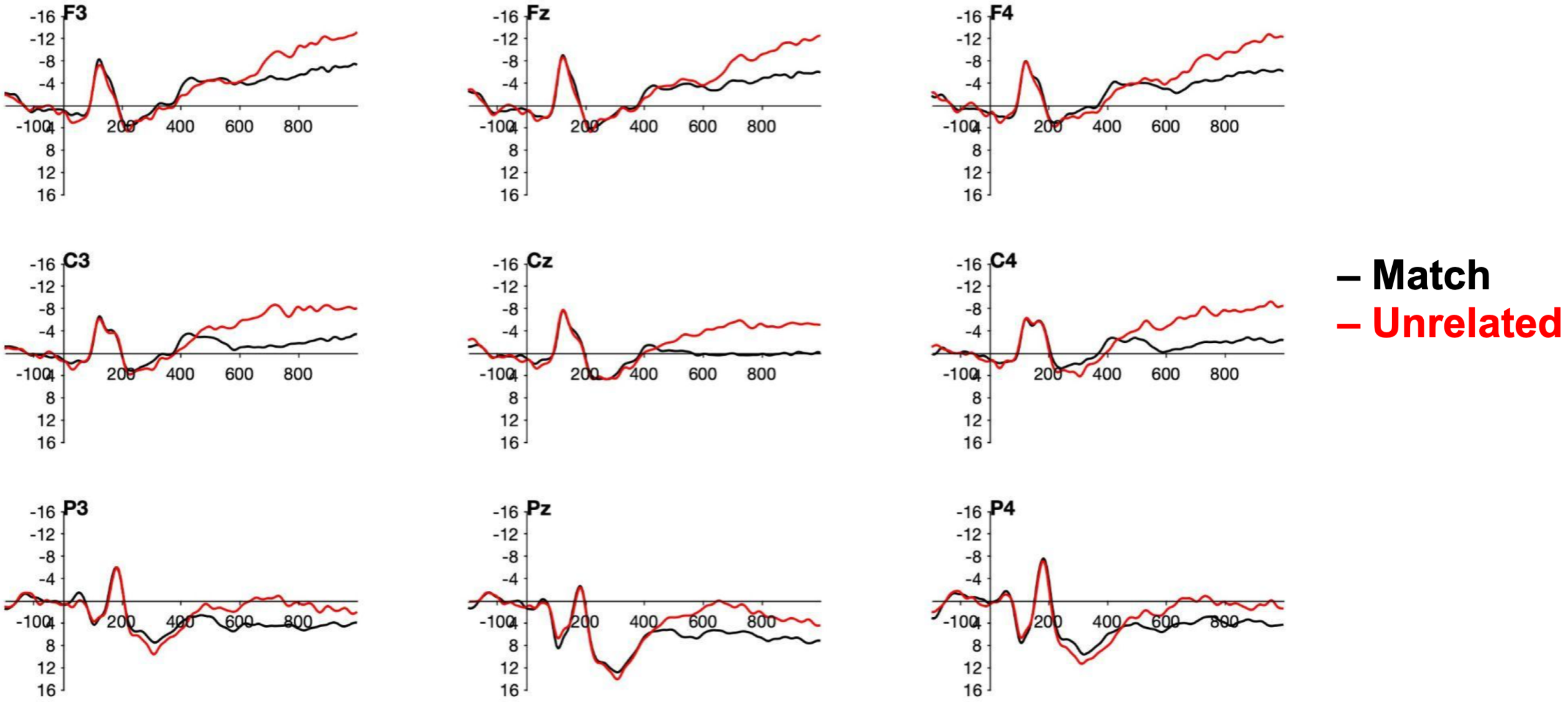
CI-Using Children (n = 13): Match Vs. Word Initial Cohort



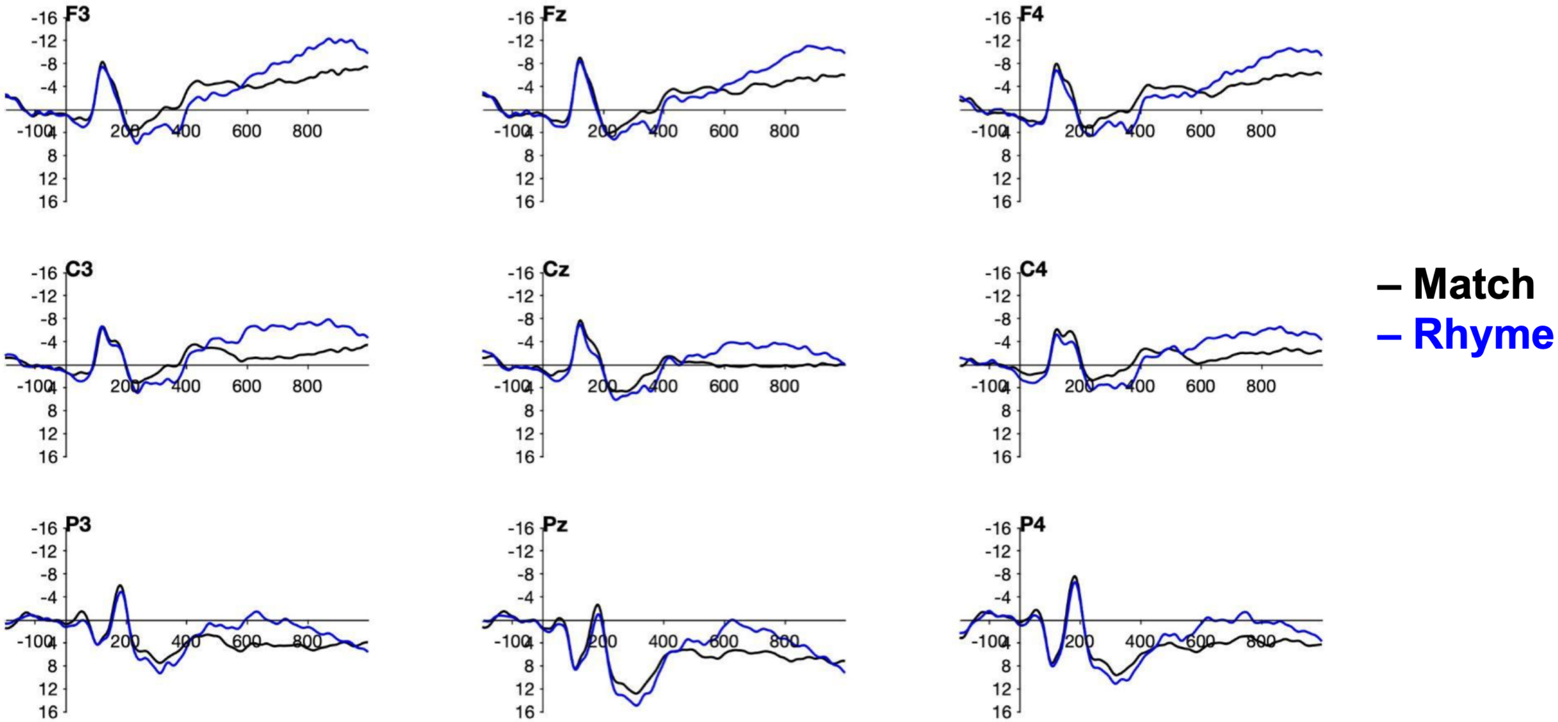
Typical Hearing Children (n = 22): All Conditions



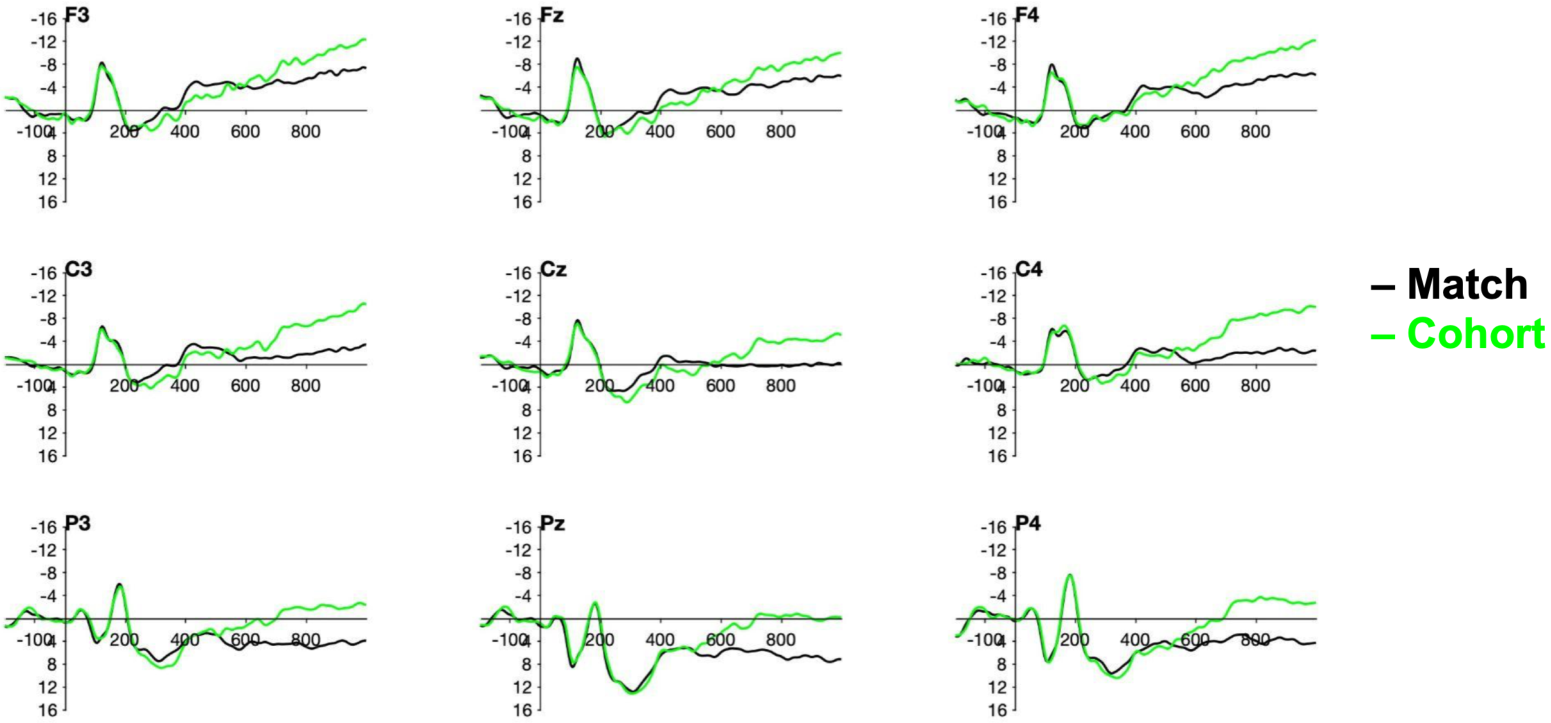
Typical Hearing Children (n = 22): Match Vs. Unrelated



Typical Hearing Children (n = 22): Match Vs. Rhyme



Typical Hearing Children (n = 22): Match Vs. Word Initial Cohort



Appendix C. Mean amplitude for component measurement windows, measured over representative site Cz and organized by group (CI-using group n = 13; TH group n = 22).

Measurement Window	200 - 350 ms (N280 / P300)		350 - 500 ms (N400)		500 - 900 ms (Late N400)	
Condition	CI	TH	CI	TH	CI	TH
Cohort	0.79 (*5.51*)	5.31 (*6.18*)	1.42 (*3.77*)	1.52 (*5.95*)	1.93 (*5.90*)	-2.38 (*6.61*)
Rhyme	0.16 (*7.10*)	5.71 (*8.28*)	0.11 (*6.34*)	0.88 (*6.66*)	0.57 (*4.87*)	-2.69 (*8.13*)
Unrelated	-0.49 (*6.62*)	4.69 (*8.56*)	–1.24 (*6.60*)	0.20 (*7.11*)	-1.29 (*7.66*)	-4.23 (*8.36*)
Match	0.96 (*4.04*)	4.01 (*5.75*)	2.09 (*3.81*)	-0.06 (*6.37*)	4.23 (*4.69*)	0.11 (*6.44*)

Note: values in parentheses represent standard deviations.

Appendix D. Pearson and Partial correlations between ERP measures and age and CI hearing factors. N = 22 Typical Hearing (TH) group; PKA = peak amplitude; TIS = Time in Sound (months); AOI = Age of Implantation (months). For the purpose of these analyses, P1 peak amplitude and latency values were calculated as the average from O1 and O2, where these measurements were taken. N1 peak amplitude and latency values were computed as an average from all nine electrodes where this component was measured (Fz, F3, F4, Cz, C3, C4, Pz, P3, and P4).

ERP Measure	TH Correlation with Chronological Age		CI Correlation with Chronological Age		CI Partial Correlation with TIS		CI Partial Correlation with AOI	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Visual P1 PKA	-.30	.17	-.07	.81	-.12	.71	.12	.71
Visual P1 Latency	**-.46**	**.03**	.46	.11	-.06	.86	.06	.86
Auditory N1 PKA	.06	.78	-.11	.73	.001	.98	-.001	.98
Auditory N1 Latency	**-.44**	**.04**	-.02	.96	-.44	.15	.44	.15

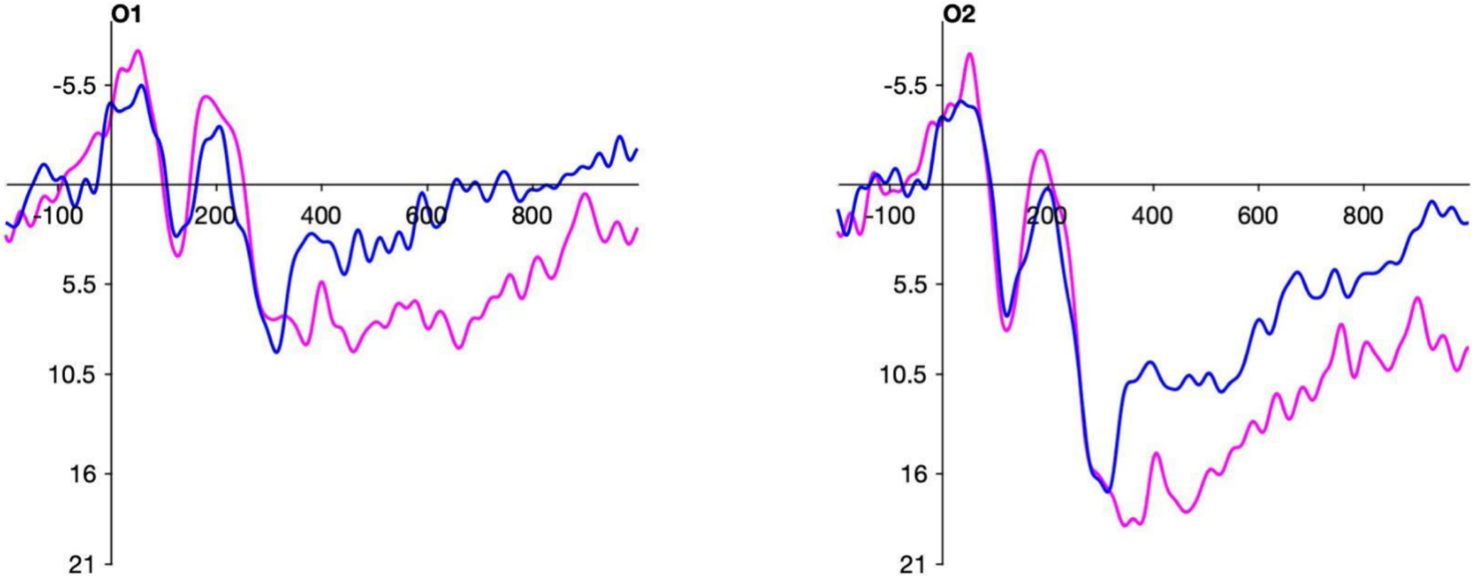
Appendix E. Correlations between ERP measures, task accuracy and response time (RT), and receptive/expressive vocabulary (ROWPVT and EOWPVT; respectively). N = 13 CI group; N = 22 Typical Hearing (TH) group; MNA = mean amplitude; N280 MNA difference values were computed as the average difference in amplitude between incongruent conditions and the match condition, from all six electrodes where this component was measured (Cz, C3, C4, Pz, P3, and P4). The Average P300 MNA for each subject was calculated as the weighted average of all conditions, at all six electrodes where the P300 was measured (Cz, C3, C4, Pz, P3, and P4). RT is calculated as the average response time for each subject on the trials that they answered correctly.

ERP Measure	Correlation with Overall Accuracy		Correlation with Overall RT		Correlation with ROWPVT		Correlation with EOWPVT	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*
CI Group								
N280 Match MNA (Baseline)	**.68**	**.01**	.25	.41	.39	.18	.54	.06
Cohort N280 Effect	-.38	.20	**-.55**	**.05**	.02	.94	.01	.98
Rhyme N280 Effect	.29	.33	-.13	.67	-.11	.71	.06	.85
Unrelated N280 Effect	-.44	.13	-.03	.92	.25	.40	.47	.10
TH Group								
Average P300 MNA	-.33	.14	-.20	.37	-.02	.93	.05	.82

Appendix F. Correlations between ERP measures and receptive/expressive vocabulary (ROWPVT and EOWPVT; respectively). N = 13 CI group; N = 22 Typical Hearing (TH) group; PKA = peak amplitude; TIS = Time in Sound (months); AOI = Age of Implantation (months). For the purpose of this partial correlation, N400 effect sizes were computed as the average difference in amplitude between incongruent conditions and the match condition, from all nine electrodes where this component was measured (Fz, F3, F4, Cz, C3, C4, Pz, P3, and P4).

ERP Measure	Correlation with ROWPVT		Correlation with EOWPVT	
	r	*p*	*r*	*p*
CI Group				
Cohort N400 Effect	-.14	.64	-.04	.90
Rhyme N400 Effect	.11	.73	.36	.22
Unrelated N400 Effect	.07	.83	.17	.58
Cohort Late N400 Effect	.02	.95	.25	.41
Rhyme Late N400 Effect	-.07	.83	-.08	.81
Unrelated Late N400 Effect	-.01	.99	.01	.99
TH Group				
Cohort N400 Effect	.10	.67	-.02	.92
Rhyme N400 Effect	.22	.32	.06	.78
Unrelated N400 Effect	-.01	.96	.05	.83
Cohort Late N400 Effect	-.13	.57	-.26	.24
Rhyme Late N400 Effect	-.22	.33	-.14	.53
Unrelated Late N400 Effect	-.13	.58	-.17	.45

Appendix G. Visual-evoked waveforms for occipital sites to Static Speaker images by group

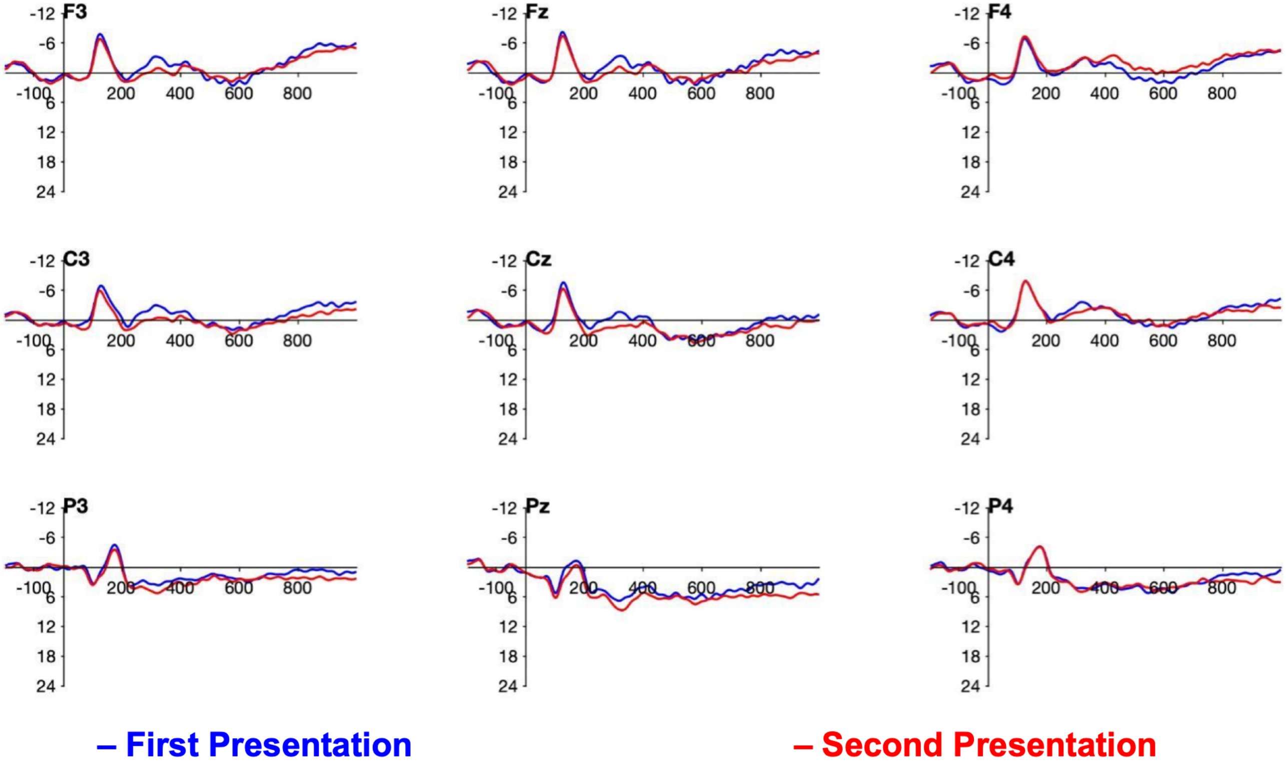


— CI-Using Group (n = 13)

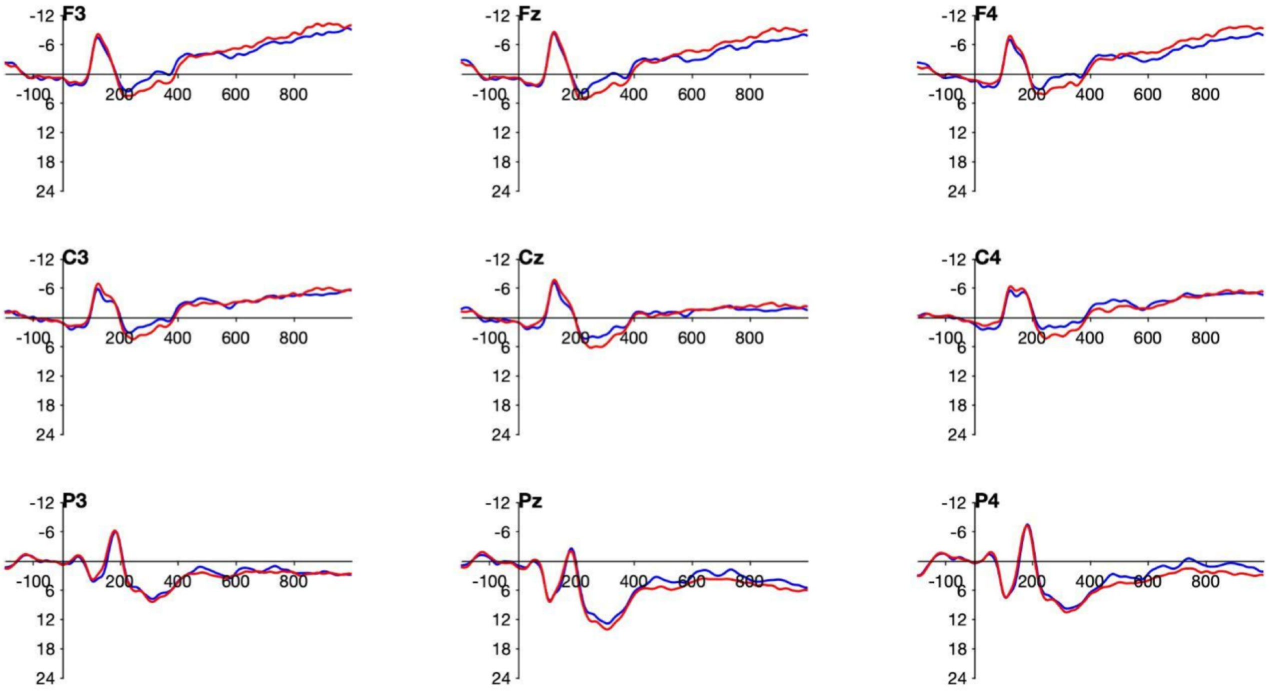
— Typical-Hearing (TH) Group (n = 22)

Appendix H. Auditory-evoked waveforms by Group and Presentation Order

CI-Using Children (n = 13): Target Presentation Order



Typical Hearing Children (n = 13): Target Presentation Order



– First Presentation

– Second Presentation