Phonetic detail, not auditory deprivation, shapes how preschoolers with cochlear implants process speech variation

Margaret Cychosz[1], Tristan Mahr[2], Benjamin Munson[3], Rochelle Newman[1], and Jan R. Edwards[1]

[1]Department of Hearing and Speech Sciences and Maryland Language Science Center, University of Maryland, College Park

[2]Waisman Center, University of Wisconsin–Madison

[3]Department of Speech-Language-Hearing Sciences, University of Minnesota, Twin Cities

Author Note

Author to whom correspondence should be addressed: Margaret Cychosz, 0100 Samuel J. LeFrak Hall, University of Maryland, College Park, College Park, USA, 20742. Email: mcychosz@umd.edu.

Abstract

To process language in real time, listeners must map a highly variable speech signal to linguistic categories such as phones and words. How do children learn to process this rampant variation and distinguish between variable pronunciations of known words ("shoup" for *soup*) versus novel words ("cheem") in their environments? By examining the unique sensory experiences of children with cochlear implants, we show that successful speech processing relies upon access to fine phonetic detail, beginning in the early preschool years. In a mispronunciation sensitivity eye-tracking task, we matched children with cochlear implants, who have poorer-resolution electric hearing, to their peers with typical acoustic hearing, controlling for (1) lifetime oral language exposure (hearing-age matches) and (2) language skill (vocabulary size matches). Online processing of familiar words (*soup*) depended on vocabulary, replicating previous work, but processing of variable pronunciations ("shoup") depended on access to fine-grained phonetic detail. These results suggest that children of all hearing experiences use their lexical knowledge to process familiar words, but they rely on fine phonetic detail and well-specified phonological representations to process variable speech in real time.

Keywords: lexical processing, speech perception, sensory experience, cochlear implant, eye tracking, hearing

1 Introduction

To develop speech and learn words, young children must learn to parse words from a highly variable speech stream spoken around them. This is a daunting task as speech is sensitive to context so word and phoneme exemplars differ from one production and speaker to the next. In absolute acoustic terms, this contextual variation—stemming from speaker pitch, rate, or accent—can be nearly as large as variation intended to be contrastive within a language. As a result, young children must disassociate *variable* word productions from *novel* word productions, in order to map truly novel words onto new referents in the environment and build a vocabulary.

Adult listeners process this speech variation easily, efficiently factoring in contextual factors in speech stemming from speaking rate, sentential context, and speaker identity or accent (Bradlow & Bent, 2008; Newman & Sawusch, 2009; Reinisch, 2016; Steffman, 2019). For example, an adult listener will consider their interlocutor's speaking rate to correctly classify temporally-based contrasts such as /k–g/ with differing voice onset times (Maslowski, Meyer, & Bosker, 2019). In this case, the listener may observe that the *absolute* voice onset time of [k] varies, but they can nevertheless classify the phoneme appropriately by computing the interlocutor's speech rate. It has long been known that because adult listeners process this systematic variation in speech so efficiently, they can quickly differentiate between variable versus novel word productions and determine if a novel pronunciation warrants a new entry in the lexicon (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978).

How do young children learn to cope with rampant speech variation that they hear in their input? When does speech variation indicate a novel accent, or a different gender, versus a new word to be learned? Without the ability to adjust for variation, children would inappropriately classify every contextual variant (e.g., [t^{h}oɪ] or [ɾoɪ] for "toy") as a novel lexical item. Thus, some degree of insensitivity is beneficial for children's speech processing. But children must likewise determine when a pronunciation variant has crossed the threshold of lexicality and a new referent must be mapped. In developmental research, these ideas have frequently been studied using "mispronunciation" sensitivity tasks, which assess how infants and toddlers process familiar words with slight ("dog" > "tog" [tɔg]) or more extreme ([sɔg]) pronunciation variants.

Over two decades of research using this mispronunciation sensitivity paradigm have demonstrated that infants and toddlers have graded sensitivity to speech variation. For instance, classic studies found that 18- to 23-month-olds looked less to a picture of a baby upon hearing "vaby" ([veɪbi]) than "baby," but still preferred the image of the baby more than a competing image of a dog (Swingley & Aslin, 2000). (See also Swingley and Aslin (2002); and Bailey and Plunkett (2002); Ballem and Plunkett (2005)). Likewise, 19-month-olds looked progressively less at the image of a ball when presented with progressively greater phonological mismatches (e.g., [gɔl] > [kɔl] > [sɔl] for "ball") (White and Morgan (2008); cf. Bailey and Plunkett (2002); Swingley and Aslin (2002)). Similar work has found that infants in this age range (15–24 months) are sensitive to vowel mispronunciations (Mani & Plunkett, 2007). And sensitivity to mispronunciations continues to be observed into the preschool years (e.g., 3–6 years (Creel, 2012)), including for lexical tone (Wewalaarachchi & Singh, 2020), with children becoming more sensitive to mispronunciations as they age (between 2;6–4;10 [years;months] (Law & Edwards, 2015)).[1]

Thus, from infancy, children are sensitive to both vocalic and consonantal mispronunciations and show graded awareness of these mispronunciations into toddlerhood. Mispronunciation sensitivity—the ability to detect, not disregard, sub-lexical variation—then appears to improve as toddlers age: children become progressively less likely to associate mispronounced words like "shoup" with the corresponding image of soup. The outstanding question is *how* children develop this perceptual flexibility that

[1] See Von Holzen and Bergmann (2021) for recent meta-analysis and Pomper, Weismer, Saffran, and Edwards (2019) for comparisons of toddlers with and without diagnoses of autism spectrum disorder.

both permits robust processing of speech variation and makes room for novel word learning. Individual difference analyses have shown that children's online word recognition of *correct* pronunciations (i.e., real words) improves with age (Fernald, Perfors, & Marchman, 2006). Studies have also demonstrated that children who have larger vocabularies and who hear more speech directed to them from caregivers in their homes process these correctly-pronounced words faster (aged 1;6–4;4: Hurtado, Marchman, and Fernald (2008); Mahr and Edwards (2018); Weisleder and Fernald (2013)). Yet vocabulary size does not reliably predict children's mispronunciation sensitivity, at least before 2;0 (Bailey & Plunkett, 2002; Swingley & Aslin, 2000; Von Holzen & Bergmann, 2021).[2] It is only between 2;0 and 3;10 that a facilitative effect of vocabulary size for mispronunciation sensitivity has been shown: children with larger vocabularies are less likely to attribute mispronunciations to familiar items (Law & Edwards, 2015; Swingley, 2016). This developmental pattern is potentially due to a restructuring of the lexicon and refinement of phonological representations with age. Indeed, vocabulary size is a stronger predictor of performance on mispronunciation tasks than age alone during this developmental stage (Law and Edwards (2015); cf. Pomper et al. (2019)).

In the current study we assess how children process speech variation by extending the mispronunciation sensitivity paradigm to a population with vastly different sensory and speech-language experience: children with cochlear implants (CIs). A CI is a prosthetic device that bypasses the middle ear to directly stimulate the cochlea and partially restore hearing for individuals with severe to profound hearing loss. The children in this study received their CI(s) between the ages of 0;6 and 3;9. Prior to implantation, they had little

[2] A recent meta-analysis found that there is no reliable effect of age or receptive vocabulary on mispronunciation sensitivity between 0;6 and 2;0, though the authors caution that the null effect of vocabulary could be due to a dearth of studies that include the measure (Von Holzen & Bergmann, 2021). There is, however, an effect of vocabulary on English-learning 16-month-olds' perception of deviant prosodic cues (Tao, Xu, & Plunkett, 2012).

to no exposure to oral language or speech input. Thus, they did not develop a vocabulary at the same pace as their peers with typical hearing (TH) because the children with CIs came from from aural/oral households and did not acquire a sign language vocabulary. Nor did the children with CIs experience spoken speech variation, stemming either from variation in their own vocal productions or variation in their caregiver's speech input (Fagan, 2014; Houston, Stewart, Moberly, Hollich, & Miyamoto, 2012).[3] Post-implantation, the children with CIs continued to have a different sensory experience than their TH peers. Because the CI stimulates the cochlea at discrete points, it breaks a continuous spectral-temporal signal (the speech envelope) into discrete components (channels), compromising the fine-grained nature of speech (for further detail see Winn and Litovsky (2015)). We refer to this hearing experience of the children with CIs as ELECTRIC HEARING to contrast it with the ACOUSTIC HEARING that the children with TH experience. Despite electric hearing's characteristic degraded signal, children who receive CIs learn to process and produce speech, hitting many speech development milestones on a similar, albeit protracted, timescale as their TH peers (Bruggeman, Millasseau, Yuen, & Demuth, 2021; Fagan, 2015; Tang, Yuen, Xu Rattanasone, Gao, & Demuth, 2019). The implant's signal itself, however, does not improve with age or cognitive development.

Thus, children with CIs in the current study have a unique sensory experience stemming from (1) the lack of oral language exposure pre-implantation and (2) the compromised speech signal post-implantation. Both sensory experiences likely shape how these children process speech and language. However, the current work focuses on the second sensory experience—the compromised speech signal—to understand how a lack of access to fine, phonetic detail shapes how children process speech variation. Specifically, it is unclear if children can rely on top-down cues from the lexicon to process speech

[3] Nevertheless, some children who receive CIs are exposed to American Sign Language or varieties of home sign. Consequently, these children have been constructing a receptive vocabulary and have been exposed to motoric production variability (in the signed modality) (Davidson, Lillo-Martin, & Chen Pichler, 2014).

variation, or, alternatively, if they must have access to fine-grained detail in their speech representations. Previous work has been unable to tease these two accounts apart because lexical growth and the specificity of speech representations are confounded in typical development. There are bidirectional influences of vocabulary and speech in development: children with larger vocabularies have more detailed, well-defined speech categories (Sosa & Stoel-Gammon, 2012; Stoel-Gammon, 2011)—potentially because children construct speech representations by generalizing over their lexicons (Edwards, Beckman, & Munson, 2004)—and children with more advanced speech (notably during the early babbling periods) go on to develop larger expressive vocabularies and language (Rescorla & Ratner, 1996; Vihman, 2014). This developmental confound does not exist for the children with CIs who experience a degraded speech signal—and have relatively underspecified sublexical phonological representations no matter how large their vocabularies become. This explanation is not meant to claim that TH children with small vocabularies have the same phonological representations as children with CIs. That is, we do not claim that small vocabularies and the degraded signal from a cochlear implant result in the same underspecified speech representations. Instead, we use the unique aural experience of children with CIs, who hear a degraded signal, to examine if lack of access to fine phonetic detail (in electric hearing) results in less sensitivity to speech variation.

In the current study, we matched children with CIs to peers with TH by language skill (vocabulary size) and lifetime oral language exposure (years of hearing experience). Matching in this way allows us to control for the children's auditory deprivation pre-implantation, as well as lexical knowledge, and isolate the effect of degraded, electric hearing post-implantation, to elucidate if and how more well-defined phonological representations help children—both with and without CIs—process speech variation in the language spoken around them.

We carried out a variant of the mispronunciation sensitivity task (Law & Edwards, 2015; Pomper et al., 2019; Swingley & Aslin, 2000; White & Morgan, 2008) where children were presented with two photos, one familiar and one unfamiliar, and heard a correct (*soup*), mispronounced ("shoup" [ʃup]), or novel word ("cheem" [tʃim]). Children's eye movements to the familiar object were then tracked. We predicted that children with CIs would distinguish less reliably between correctly-pronounced (*soup*) and mispronounced words ("shoup") than TH controls (i.e., children with CIs would look more to an image of soup upon hearing "shoup") owing to the degraded speech signal generated by electric hearing. More specifically, since the children with CIs were matched to TH controls for vocabulary, we predicted that they would process correct pronunciations (*soup*) at a similar timescale as TH controls, and that group differences would instead stem directly from the children with CIs' insensitivity to the mispronunciations ("shoup"). Together, these results would suggest that sensitivity to phonological variation during online processing in the preschool years depends on access to fine, phonetic detail and well-specific phonological representations.

Finally, previous research suggests that vocabulary facilitates the online processing of both correctly produced and mispronounced words for children with TH within this age range (Law & Edwards, 2015). Consequently, we anticipate positive correlations between the magnitude of children's sensitivity to mispronunciations and vocabulary size for both children with TH and CIs. However, categorical speech perception is likewise linked to phonetic skill in children of this age range (Rvachew, n.d.; Shiller, Rvachew, & Brosseau-Lapré, n.d.). Furthermore, since the current study primarily predicts that group differences in mispronunciation sensitivity may stem primarily from phonetic effects, not lexical (since the groups are matched for vocabulary), we additionally anticipate positive correlations between performance on the mispronunciation task and a standardized measure of the children's phonetic skill (articulatory acumen).

2 Methods

2.1 Participants

Data in this study came from N=37 observations of children with CIs and N=37 children with NH. Data from an additional 4 children with CIs were collected but were excluded due to missing eye gaze data (see Section 2.4). All children were monolingual English speakers and were participating in a larger longitudinal research program where children's vocabulary and phonological development were assessed annually at age 3, 4, and 5.

The children with CIs were matched to the children with NH for hearing age, vocabulary size, gender, and maternal education using the R package `Matching` (Sekhon, 2011) (see Table 1 for further detail). After controlling for all relevant variables, we were able to make N=19 matches. We report separate analyses for both the matched pairs (N=19) and all observations of children with CIs, including those that were not matched (N=37).

Maternal education was reported by the children's caregiver(s) and was divided into seven levels for matching: (1) < high school degree, (2) equivalent of high school degree (GED), (3) high school degree, (4) technical-associate's degree, (5) some college, (6) college degree, and (7) graduate degree. To facilitate matching, while still ensuring a sufficient sample size, N=8 children (N=4 with CIs, and N=5 with TH) contributed data from 2 out of the 3 timepoints that they were observed (e.g., at age 3 and 4). For clarity, we refer to these repeated observations as *unique* children throughout. We explain the statistical modeling of these repeated observations in the results.

All of the children with TH had typical speech and language development, per parental report. The children with TH additionally all passed a standard hearing test in at least one ear at 25 dB at 1000, 2000, and 4000 Hz. The children with CIs had profound

Table 1

Participant demographic information (N=19 matches). To facilitate matching, N=4 children with CIs and N=5 children TH were observed at two timepoints (e.g. at age 3 and 4). Mean (SD), range.

Hearing Status	Girls, Boys	Chronological Age: months	Hearing Age: months	Maternal Education Level	EVT-2 GSVs	EVT-2 Stan. Scores
Children with CIs	13, 6	56.3 (6.7) 44–66	44.7 (8.7) 29–56	6.26 (0.56)	134.84 (12) 112–159	102.63 (13.37) 84–131
Children with TH	13, 6	Not applicable	45.5 (7.5) 36–57	6.05 (0.62)	132.68 (10.88) 117–150	114.58 (10.38) 98–134

deafness in both ears. N=21 had bilateral CIs, N=3 had 1 hearing aid and 1 CI, and N=1 child had 1 CI. The average age of CI activation was 18 months (SD=10.6; range=6–45).[4] The children with CIs completed the Ling6 sound test prior to experimental testing to ensure CI functioning. Additional audiological information is included in the Appendix.

2.2 Stimuli

Lexical stimuli were N=6 easily-picturable, one-syllable CVC nouns, familiar to 90% of 30-month-olds according to MacArthur-Bates Communicative Development norms (Fenson et al., 2007). The corresponding mispronunciations were created by manipulating one feature of the initial consonant of each word (e.g., *soup* > [ʃup]). The N=6 novel words were also CVC and were phonotactically matched to the real words on the basis of transitional probabilities between each C-V and V-C using the Hoosier Mental Lexicon (Pisoni, Nusbaum, Luce, & Slowiaczek, 1985). See Table 2. We do not model looking patterns in response to the novel words for the remainder of this work.

[4] These statistics refer to all 25 unique children with CIs who were analyzed in the current study, not just the 15 unique children with CIs who were age- and language-matched.

Table 2

Lexical stimuli and manipulations used in eye-tracking paradigm.

Contrast and Manipulation	**Target Word**	**Mispronunciation Transcription**	**Novel Word**
/s/ > [ʃ]	soup	[ʃup]	cheem [ʧim]
/ʃ/ > [s]	shoes	[suz]	geev [giv]
/g/ > [d]	girl	[dɝl]	shan [ʃæn]
/d/ > [g]	duck	[gʌk]	neydge [neɪʤ]
/k/ > [g]	cake	[gek]	pum [pʌm]
/r/ > [w]	rice	[waɪs]	bape [bep]

Visual stimuli consisted of pairs of color photos, consisting of one familiar item (e.g., soup) and one item unfamiliar to these children (e.g., a sextant). The pairings remained consistent throughout the study; for example, images of soup (familiar) and bamboo steamers (unfamiliar) always appeared together. The left-right position of familiar versus unfamiliar photos on the screen was counterbalanced between trials. To maintain children's attention, two different photos of each item were used on different trials. For the novel word trials, children were likewise presented with one familiar (unrelated to the six target familiar items) and one unfamiliar item. These items were likewise familiar to at least 90% of 30-month-olds. Photos were matched within and between trials for size, animacy, and attractiveness, per the authors' judgment. Photos were normed by N=30 children from two preschools; see supplementary material I for details.

Auditory stimuli consisted of recordings of each lexical item made by a female adult in a child-directed speech register spoken in the local dialect. Lexical items were embedded in carrier sentences such as "Find the *X*!" and "See the *Y*!" To mitigate any effect of coarticulatory cues in these carrier sentences, neutral sentences with lexical items beginning with a glottal stop, to facilitate cross-splicing, were also recorded (e.g., VC in

"Find the *egg*!"). Then, the target lexical item was appended to the neutral carrier phrase ("Find the$_{\text{[neutral]}}$ soup$_{\text{[target]}}$.") with 80 ms between the carrier phrase and target item.

Duration was normalized between familiar items and their corresponding mispronunciations and novel words (e.g., familiar *soup* > mispronunciation [ʃup] > novel word [ʧim]). Intensity was normalized between all items.

2.3 Task procedure

A looking-while-listening procedure was conducted (Fernald, Zangl, Portillo, & Marchman, 2008). Children were seated approximately 60 cm in front of a monitor screen and guided by an experimenter in the room through the task. Eye gaze was recorded using a Tobii T60XL eyetracker (60 Hz sampling rate, though this was downsampled to 50-ms windows for a rate of 20 Hz to smooth over data from adjacent frames). For each trial, photos of a familiar and unfamiliar object were centered side-by-side on a gray background. Auditory stimuli, presented at approximately 65 dB, were played from a speaker under the monitor.

Trials fell into one of three different conditions. For Correct Pronunciation trials, photos of the target familiar object and an unfamiliar object were presented with recordings of the correct pronunciation of the familiar object (e.g., *soup*). Mispronunciation trials were identical except that the auditory stimuli were one-feature mispronunciations of the target familiar object. Novel Word trials were presented with different pairs of familiar-unfamiliar photos and the accompanying recording of the novel word. Trials were interspersed with 500 ms of a blank screen. Before beginning the experiment, the eyetracker was five-point calibrated to each child. After approximately every 6 trials, an attention getter was played and the experimenter ensured that the child was still sitting sufficiently close to the tracker and that their activity was being tracked.

Each child completed two experimental blocks. Each experimental block contained

12 trials per condition (Correct Pronunciation, Mispronunciation, and Novel Word) for 36 trials per block. Trials were pseudo-randomized such that (1) each block began with a Correct Pronunciation trial, (2) no more than two trials of a given condition were presented in a sequence, and (3) the correct pronunciation was never presented with its accompanying mispronunciation in the same block of trials. Each child completed 2 blocks. Children completed a standardized screening between blocks.

Each trial unfolded as follows: both photos were presented in silence for 2,000 ms, to familiarize the child. Then, presentation of the auditory stimulus was gaze-contingent: the experimental software attempted to continuously track the child's eye gaze movements. After 10 seconds, if the child's eye gaze movements had not been continuously tracked for 300 ms, the trial continued. After eye gaze verification, the carrier phrase and target word played. This procedure attempted to guarantee that the child was looking at the screen when the audio stimulus played. Then, 1,000 ms after target offset, a reinforcer phrase such as *You're doing great!* played and the images remained on the screen for another 1,000 ms. Additionally, reinforcer images and phrases were presented every 6–8 trials to maintain children's interest.[5]

2.4 Data cleaning

Before data cleaning, we performed "deblinking" to account for gaze patterns that were lost due to blinking and not, for example, looking off screen. Short windows of data (up to 150 ms) were interpolated if the child fixated on the same image before and after a missing data window. At the trial level, data quality was examined in the 250–1,500 ms window following target word onset: at least 50% of the data within the window had to be valid (onscreen) to include the trial (CIs: N=651 trials removed; TH: N=370). Next, at the block level, at least N=12 trials within the block had to be valid to include the block

[5] The reinforcer phrase was dropped for the observations made at age 5 to accommodate the older children.

(additional 6 blocks removed). Finally, at the condition level, at least N=6 trials had to include valid data to include the condition (0 conditions removed). On the basis of these criteria, we removed 4 children with CIs from analysis entirely. The remaining N=37 children with CIs were matched to a subset of TH controls (selected semi-randomly from the approximately N=160 children with TH who completed the tasks). N=425 trials remained from the N=19 matches that we made (CIs: N=212 trials; TH: N=213 trials, out of a total possible 228 trials).[6]

2.5 Speech-language measures

To assess correlations between children's vocabulary size, articulation and their mispronunciation sensitivity, we additionally had the children complete standardized tests of vocabulary and articulation. Vocabulary was assessed with the Expressive Vocabulary Test, 2nd edition (EVT-2) Williams (2007) and consonant articulation skill was assessed with the Sounds-in-Words portion of the Goldman-Fristoe Test of Articulation, 2nd edition (GFTA-2) Goldman and Fristoe (2000). These speech-language measures are included to explain individual differences between children within each hearing group (i.e., not to compare children with CIs and TH). Consequently, we include data from *all* children who completed the vocabulary and consonant articulation tasks (N=33/37 children with CIs, and N=24/37 children with TH) and not just the children who were matched for gender, maternal education, etc.[7] See Table 5 for descriptive statistics of assessment results by hearing status and Supplementary Materials I for demographic information for these children.

For the vocabulary test, children were presented with an image and asked to name it

[6] N=7 children with TH heard an alternation of *dog* - "tog" instead of *rice* - "wice"; these trials were additionally removed.

[7] Fewer children completed both assessments because the articulation task was not measured for the children with TH at age 5.

Table 3

Summary statistics of standardized speech-language measures for ***all*** *children, by hearing status (N=33 children with CIs and N=24 with TH). Mean (SD), range. Scores are used to assess individual differences within groups and as such, all children are included and children are not matched by hearing experience, gender, SES, or language.*

Hearing Status	EVT-2 standard score	EVT-2 GSVs	GFTA-2 standard score
Children with CIs	95.7 (18.71) 46–127	120.76 (25.99) 42–159	73.61 (19.33) 39-107
Children with TH	116.17 (12) 88–134	140.46 (15.46) 117–164	90.04 (12.04) 67-113

or provide a synonym. For the articulation test, children were asked to repeat N=53 picture-prompted words. Children's productions were audio recorded for offline scoring. N=37 singleton consonants in word-initial, -medial, and -final positions (onset and coda), and N=16 consonant clusters in word-initial position, were then scored. Only omissions, substitutions, and non-responses, but not epenthesized segments, were marked incorrect. Our statistical modeling includes growth scale values for vocabulary (transformations of raw scores that grow linearly with age) and standard scores for articulation skill (scores normalized for sex and age).

3 Results

We evaluate the mispronunciation sensitivity of children with CIs in comparison to their hearing age- and vocabulary size-matched peers. The outcome variable is the proportion of looks to the familiar object versus the unfamiliar object as a function of time (300–1,800 ms after target word onset). We modeled these looking proportions using Generalized Additive Mixed Models (GAMMs). This approach has become an important tool to model time series data, such as eyetracking trajectories, because it can estimate flexible, nonlinear relationships ("smooths") between variables such as time and relevant covariates (i.e., effects of group and/or condition) (Rij, Hollebr, & Hendriks, 2016; Zahner, Kutscheid, & Braun, 2019). GAMMs are composed of (fixed) *parametric* terms that model static relationships between two variables, as is common in generalized linear modeling, and *smooth* terms that model nonlinear effects by using penalized basis functions (i.e., smoothing splines). Parametric terms can typically be interpreted from model summaries, as in traditional regression, but smooth terms must be interpreted visually. Wieling (2018) and Sóskuthy (2021) provide tutorials for GAMMs in linguistics, and Wood (2017a) provides a comprehensive textbook treatment of the approach.

GAMMs also allow for autocorrelation between observations to be factored into the modeling. Incorporating autocorrelation is of particular importance to eyetracking data where we anticipate large amounts of within-trial correlation between measurements over time: the area where the child is looking at 500 ms is highly correlated with where they are looking at 550 ms. As such, GAMMs are a significant improvement upon other polynomial regression models common in time series analysis, such as Growth Curve Models (GCMs). The standard approach for estimating GCMs cannot factor in this inherent correlational structure within the data and, as a result, recent work has shown that they result in inflated Type I error rates (Huang & Snedeker, 2020).

The current data were analyzed in the RStudio computing environment (R version 4.0.2; R Core Team (2020)). All computing and statistical analyses are included in the GitHub repository affiliated with this project (`github.com/megseekosh/ci-mispron`). Visualizations were made using the `ggplot2` (Wickham, 2016) and `cowplot` (Wilke, 2020) packages. Modeling was conducted and presented using the `mgcv` (Wood, 2017b), `itsadug` (van Rij, Wieling, Baayen, & van Rijn, 2020), and `tidymv` packages (Coretta, 2022) (see project documentation for package versions). For all modeling, the proportion of children's looks to the familiar object versus the unfamiliar object was calculated for each frame (every 50 ms) and transformed to empirical logit (elog), or the log-odds of looking to the familiar object at each sample (Barr, 2008). Random effects ("factor smooths" in GAMMs) included by-participant, by-observation (first [at age 3], second [at age 4], or third [at age 5] visit to the lab), and by-item trajectories. These factor smooths modeled variability stemming from individual children and lexical items, and took into account the repeated observations from some children at two different ages. To ensure assumptions were met and to avoid overfitting, model criticism was conducted using the `gam.check()` function; when necessary, the number of basis functions (*k* or knots) was increased.

As is common in eye-tracking data, the response data were distributed with heavy tails. Consequently, all models were fit using a scaled-*t* model using the `scat()` link function, which substantially improved data distribution (Wood, Pya, & Saefken, 2016). Finally, for each model, autocorrelation between model residuals was calculated; all models showed high amounts of autocorrelation. These dependencies were factored into the modeling by allowing AR(1) an autoregressive error parameter that modeled the degree of autocorrelation (*rho*) between timepoints in each trial. Subsequent model inspection demonstrated that specifying this autocorrelation value in the model sufficiently factored out autocorrelation between residuals (Wieling, 2018).

3.1 Evaluating the effect of phonetic detail on mispronunciation sensitivity

To evaluate how access to fine, phonetic detail may affect mispronunciation sensitivity, a series of GAMMs were fit comparing children with CIs and their hearing age- and vocabulary size-matched TH peers. **Condition** (Correct Pronunciation vs. Mispronunciation) was contrast-coded to facilitate model interpretation and the 2x2 relationship of **Group** (Children with CIs vs. TH) and **Condition** was modeled using ordered factors. A model with parametric and smooth terms for **Group** and **Condition** improved upon a **Condition**-only model, suggesting that children with CIs and TH responded differently to correct pronunciations versus mispronunciations.

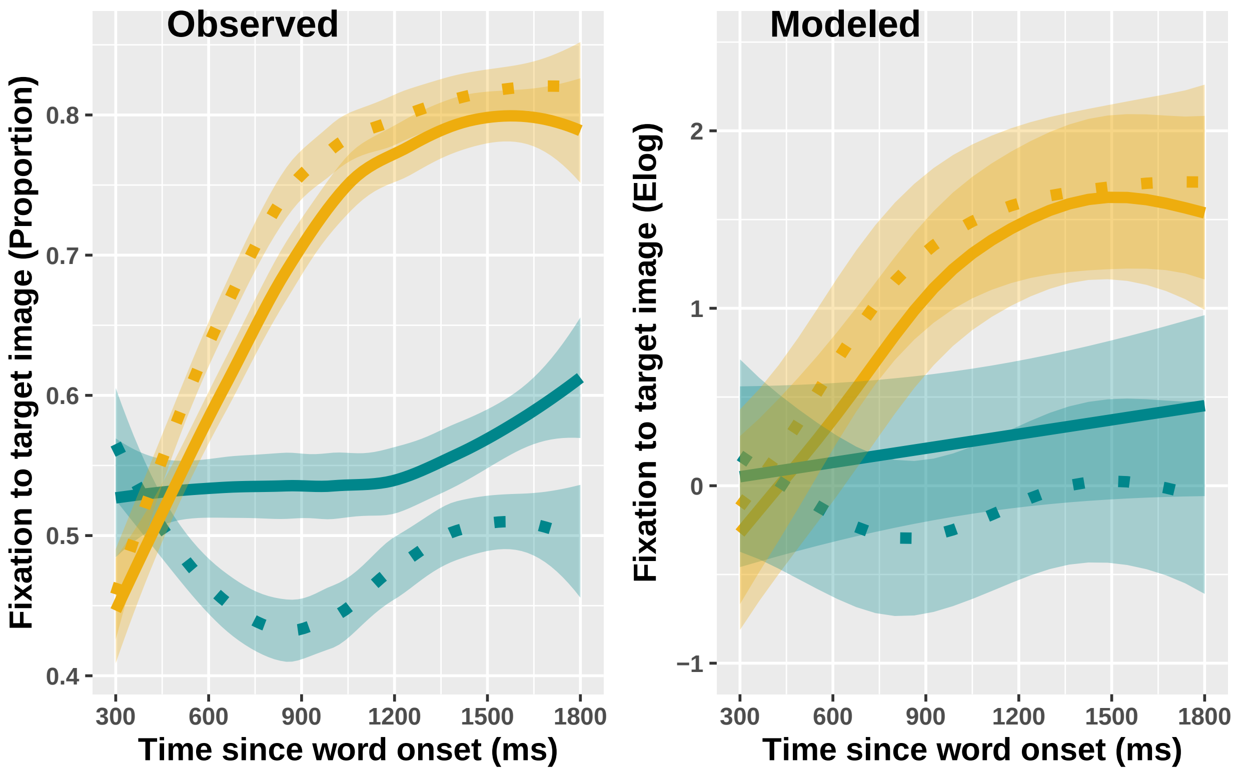
To statistically evaluate the source of the **Group** effect (i.e., stemming from overall vs. time-varying response to the stimuli), another model was fit that included parametric terms for **Group**, and the ordered factors of *Correct Pronunciation* for children with CIs and *Correct Pronunciation* for children with TH (Wieling, 2018). These parametric effects modeled the constant effect of the covariates upon the response variable; smooth terms are centered around 0, and these parametric effects adjust these curve to center at some average proportion of looks. Smooth model terms included nonlinear effects of **Time** and **Time** by **Group**. The latter allowed us to model the nonlinear difference between the two different groups' responses to mispronunciations. Finally, the model included difference smooths, which allowed us to separately model how each hearing group responded to correct- versus mis-pronunciations, over time. See Table 4 for model summary.

In the first part of the results, we ask: are both children with CIs and their TH matches sensitive to mispronunciations? Parametric effects in the model summary show that there are, overall, significantly more looks to the familiar photo for *Correct Pronunciation* trials than *Mispronunciation* trials, for both children with CIs and TH (CI logit Est.=0.84, p<.001, proportion Est:0.21; TH logit Est.=1.31, p<.001, proportion Est:0.31). We interpret the smooth terms by first considering effective degrees of freedom (EDF) and the significance test for each smooth. The EDF indicates how much wiggliness there is in a smooth where EDF=1 indicates a linear relationship and a larger value indicates more wiggliness in the smooth. Interpretation of the nonlinear smooths shows that there are significant, nonlinear differences in looks to the familiar object between correct- and mis-pronunciations for children with CIs (smooth of **Time** by the ordered **Cochlear Implant; Correct**) and children with TH (smooth of **Time** by the ordered **Typical Hearing; Correct**) (Figure 1). Thus, both children with CIs and TH are sensitive to mispronunciations.

Parametric coefficients	**Estimate**	**Std. Error**	**t-value**	**p-value**
Intercept (Cochlear Implant: Mispronunciation)	0.27	0.19	1.39	0.17
Typical Hearing	-0.33	0.19	-1.69	0.09
Typical Hearing: Correct	1.18	0.23	5.24	< 0.001
Cochlear Implant: Correct	0.68	0.23	3.01	0.003
Smooth terms	**edf**	**Ref. df**	**F-value**	**p-value**
s(Time)	0.00	0.00	755.41	0.58
s(Time,Cochlear Implant)	1.00	1.00	2.63	0.11
s(Time,Typical Hearing)	3.26	4.14	1.36	0.24
s(Time,Typical Hearing: Correct)	5.06	6.32	8.65	< 0.001
s(Time,Cochlear Implant: Correct)	3.76	4.83	7.28	< 0.001
s(Time,Child)	52.18	341.00	0.47	< 0.001
s(Time,Item)	25.54	108.00	0.69	< 0.001
s(Time,Observation)	0.00	27.00	0.14	< 0.001

Table 4

Model summary predicting the difference between proportion of looks to the familiar object by word condition and hearing status.



Nevertheless, the above modeling cannot tell us if these children with CIs are *less* sensitive to mispronunciations than their TH peers; the modeling demonstrates only that both groups show sensitivity. To evaluate differences in mispronunciation sensitivity by group, another GAMM was fit, with a binary difference smooth, which allowed us to evaluate the *difference* between smooths (Correct- vs. Mis-pronunciations) for children with CIs and TH, over time. Model fit included parametric effects of **Group**, as well as smooths of **Time**, **Time** by **Group**, **Time** by **Condition**, and **Time** by the ordered variable of **Group** by **Condition** (to model the difference between real- and mis-pronunciations for each group). Model results are plotted in Figure 2; the model summary is included in supplementary materials. Overall, the model-estimated difference smooths show smaller differences between correct- and mis-pronunciations for the children with CIs—and that these differences take longer to manifest during online processing (left panel of Figure 2). Further inspection of the first model, as plotted in Figure 3, demonstrates why this is the case. The children with CIs and TH do not respond significantly differently to correct pronunciations: once vocabulary size and hearing age are controlled, both groups of children respond similarly to correctly-pronounced words. Instead, children with CIs—who are listening with a degraded speech signal via electric hearing—are less sensitive to *mis*-pronunciations (Figure 3), resulting in smaller difference smooths between correct- and mis-pronunciations.

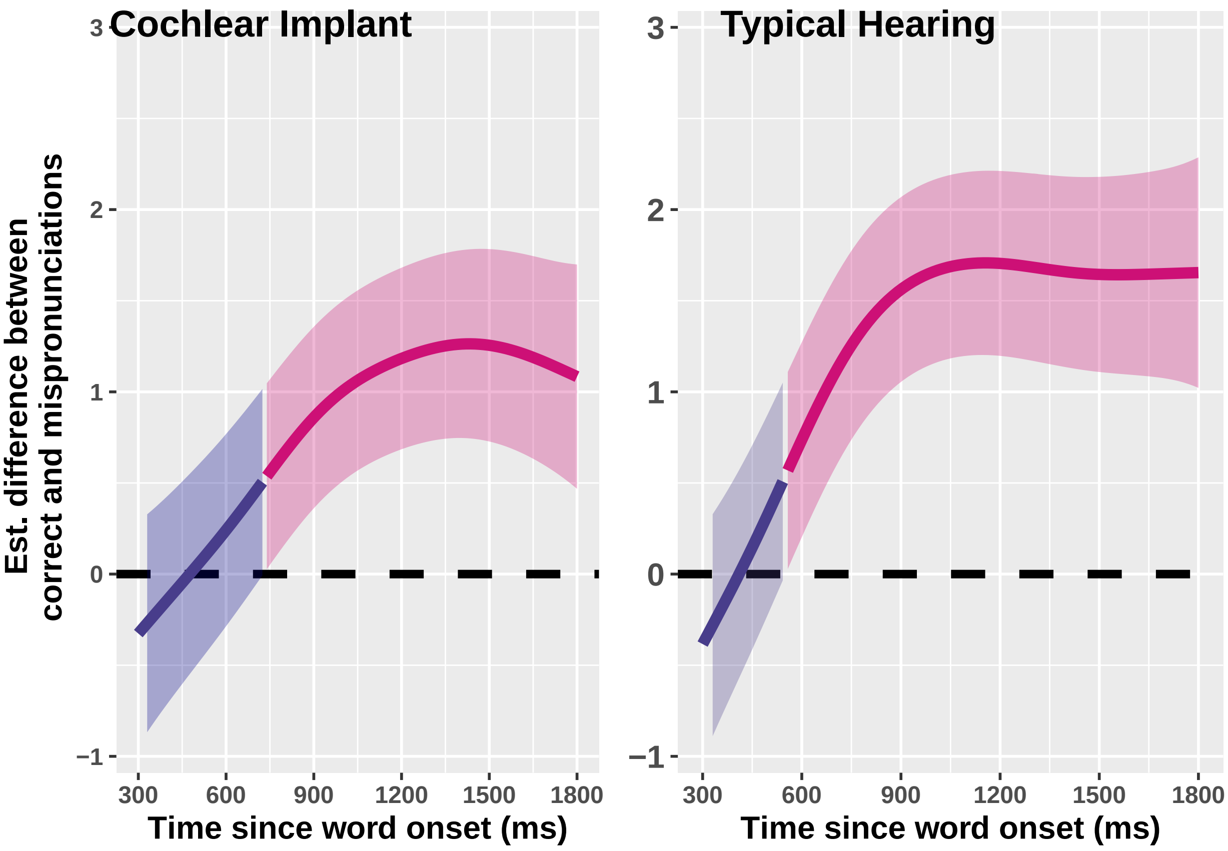


Figure 2. Difference smooths (GAMM predictions) by condition (correct- vs. mis-pronunciations) for children with CIs (L) and TH (R). Pink smooths represent the point when correct- and mis-pronunciation smooths differ (i.e., reliable effect of condition) for each group: there is a larger difference between correct- and mis-pronunciation responses for children with TH than CIs. Shaded ribbons represent 95% confidence intervals.

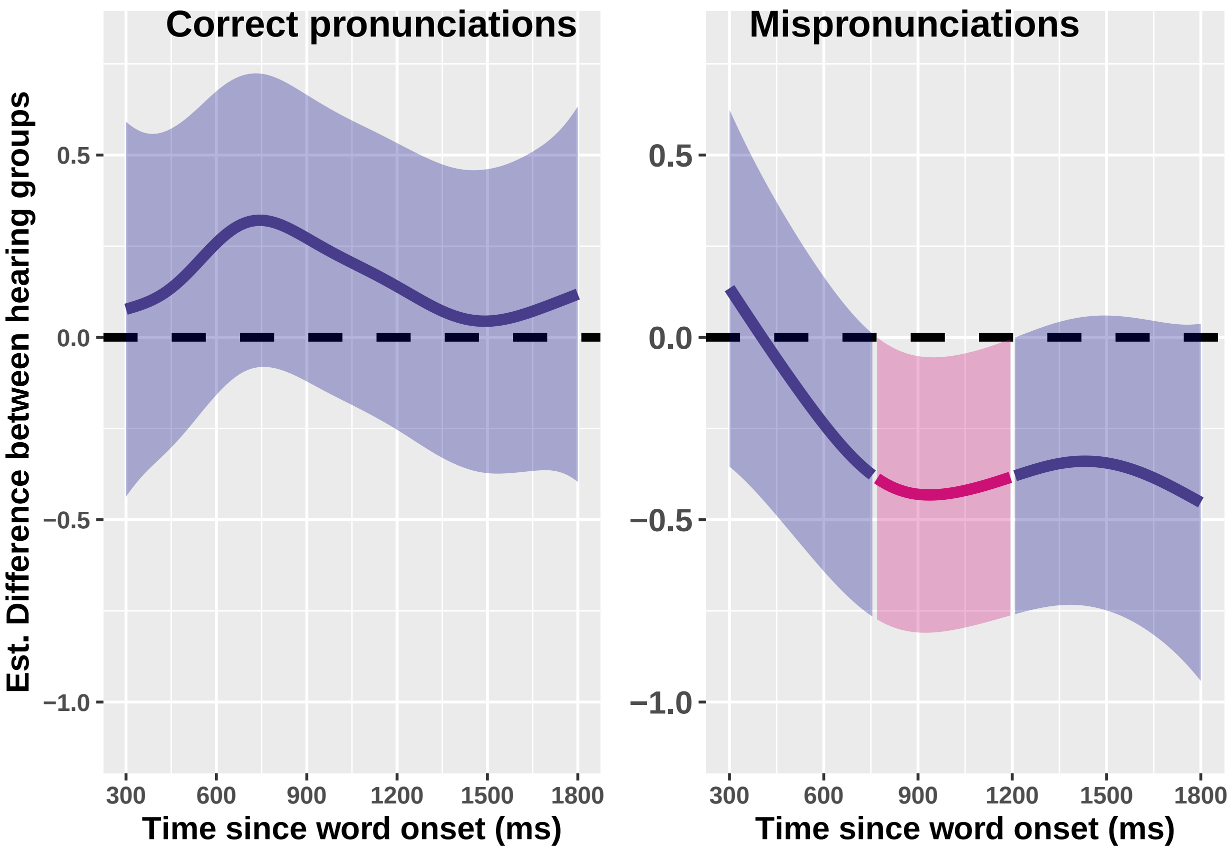


Figure 3. Difference smooths (GAMM predictions) by hearing status for correct pronunciations (L) and mispronunciations (R). The pink smooth represents the point when the smooth for children with CIs differs from children with TH (i.e., reliable effect of group): there is an effect of group upon mispronunciations, but not correct pronunciations. Shaded ribbons represent 95% confidence intervals.

3.2 Explaining individual differences in mispronunciation sensitivity

Having established that children with CIs are less sensitive to mispronunciations than their TH peers, we next correlated the children’s responses with two different standardized speech-language assessments: expressive vocabulary size (EVT-2) and spoken phonetic/articulatory accuracy (GFTA-2). Because we took an individual differences approach, we examined the children with CIs and TH separately.

We modeled the effects of vocabulary size and articulation on the children who

Table 5

Summary statistics of standardized speech-language measures, by hearing status (N=33 children with CIs and N=24 with TH). Mean (SD), range.

Hearing Status	EVT-2 standard score	EVT-2 GSVs	GFTA-2 standard score
Cochlear Implant	95.7 (18.71) 46-127	120.76 (25.99) 42-159	73.61 (19.33) 39-107
Typical Hearing	116.17 (12) 88-134	140.46 (15.46) 117-164	90.04 (12.04) 67-113

completed both assessments (N=33 with CIs and N=24 with TH) by using stepwise GAMM fitting. Specifically, we assessed the non-linear interaction between **Time**, **Condition**, and **Vocabulary Score/Phonetic Accuracy** to evaluate if children's vocabulary sizes and/or phonetic accuracy predicted their looks to the target over time for the correct- and mis-pronunciation conditions. As before, all models included factor (random) smooths by participant, observation (visit to the lab), and item. Each additionally included a difference smooth of **Time** and **Participant** by **Condition** (*Correct-* versus *Mis-pronunciation*). A baseline model was fit with a parametric term for **Condition** (estimating the average looking probability in each condition), smooth terms for **Time** and **Time** by **Condition**, as well as a non-linear interaction (tensor product) of **Time** and **Child Age** by **Condition**. In all models we included the Age by Condition tensor product smooth term because our child-level variables (vocabulary score and phonetic accuracy) are confounded with age, and we wanted to evaluate the potential influence of these speech-language abilities independent of child age.

We fit the three-way smooth interaction of **Time**, **Condition**, and **Vocabulary Score** and **Time**, **Condition**, and **Phonetic Accuracy** using tensor product terms. For the children with TH, neither the vocabulary nor phonetic accuracy term improved upon a baseline model controlling for child age. This result indicates that, for the children with TH, mispronunciation sensitivity—the difference in looks to the target image in correct- versus mis-pronunciation conditions—is not moderated by vocabulary size or phonetic accuracy over and above age effects. For the children with CIs, best model fit included **Phonetic Accuracy**; **Vocabulary Score** did not improve upon model fit. The final model summary for the children with CIs is included in the supplementary materials.

Given the multiple non-linear effects at play, it is necessary to plot the model predictions in order to interpret GAMM outputs, in particular how phonetic accuracy mediates mispronunciation sensitivity for the children with CIs. To facilitate interpretation of the non-linear three-way interaction, the children with CIs were divided into tertiles by vocabulary score and phonetic accuracy. Predictions from the model, by articulatory tertile, are plotted in Figure 4 and raw response curves are plotted in Figure 5. The model predictions demonstrate that children with better articulation scores show larger differences between looks to the target for correct- versus mis-pronunciations (higher overall *y*-intercept value) and that these children show significant differences between correct- and mis-pronunciations slightly earlier in the analysis window (cross-over from purple to pink smooth occurs sooner in the analysis window). Thus, for the children with CIs, phonetic/articulatory accuracy predicts mispronunciation sensitivity, independent of age and language ability.

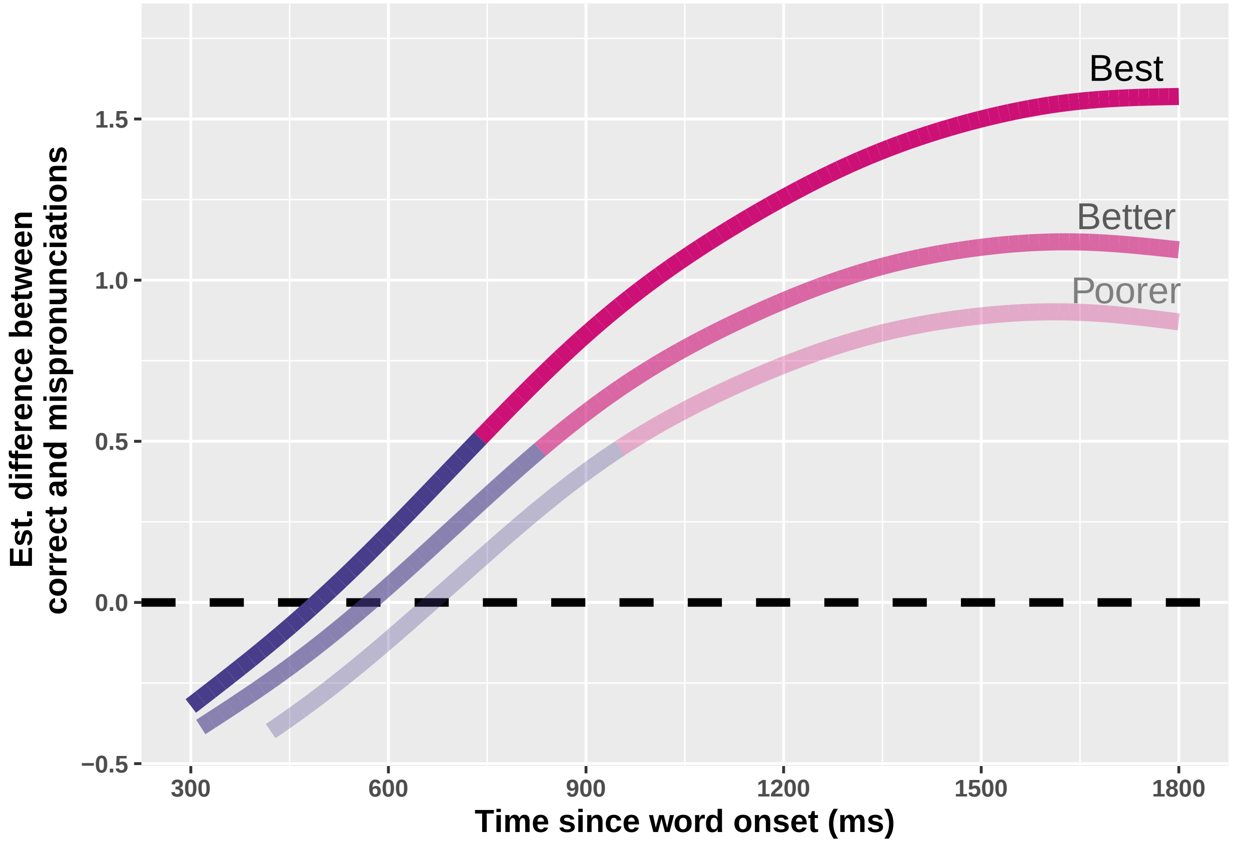


Figure 4. Difference smooths (GAMM predictions) between correct- and mis-pronunciations for children with CIs, by standardized articulation score. Pink smooths represent the point when correct- and mis-pronunciations smooths significantly differ (i.e., reliable effect of condition). Children were divided into tertiles by score with smooths representing the median score for children with poorer (median score=57), better (72), and best (96) articulation scores.

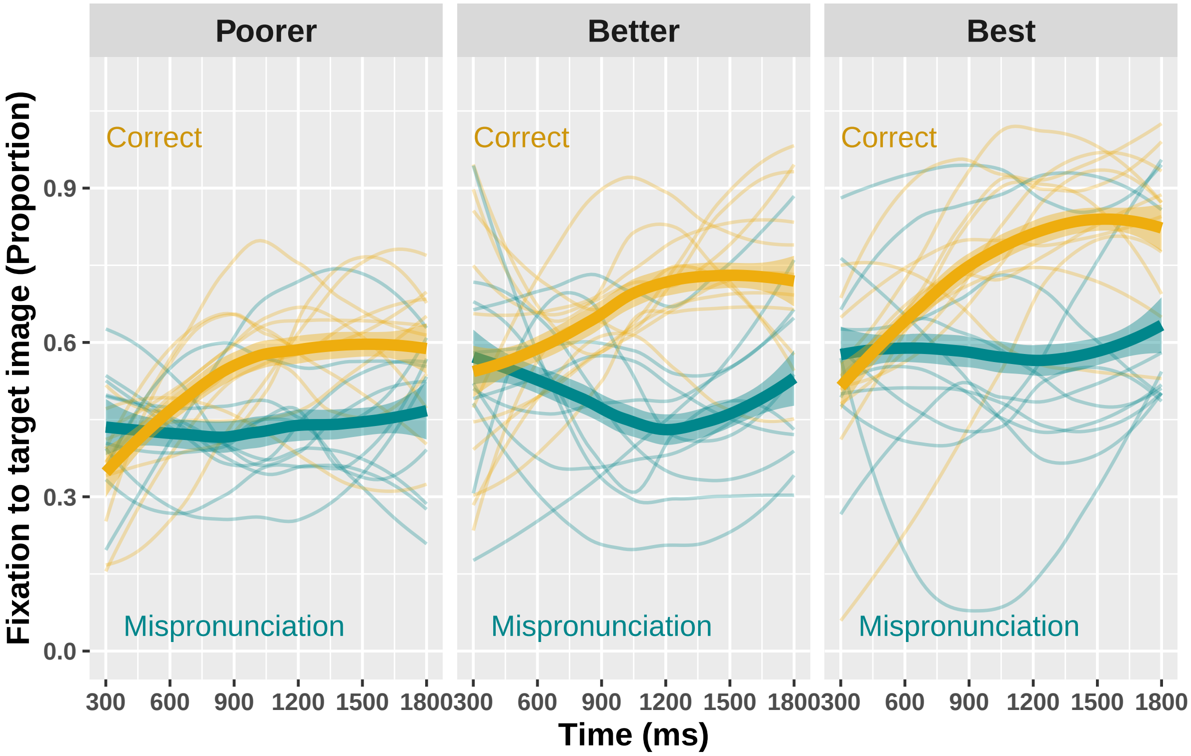


Figure 5. Raw response trajectories for proportion of looks to familiar object for children with CIs, by word condition and standardized articulation score. Children were divided into tertiles by score: poorer (median score=57), better (72), and best (96) articulation scores.

4 Discussion

This study asked how preschoolers learn to process variation in speech by taking advantage of the unique sensory experiences of cochlear implant (CI) users. In the past, it has been unclear if children relied on top-down cues from the lexicon to process speech variation, or if they instead must have access to fine-grained representations. The unique sensory profile of CI users allows us to disentangle these two explanations. The electric hearing generated by the CI results in a degraded speech signal which allowed us to assess how (lack of) regular access to fine, phonetic detail affected preschoolers' speech processing, while controlling for lexical knowledge (vocabulary size).

We carried out a variant of the mispronunciation paradigm where children responded to correct- (*soup*) and mis-pronunciations ("shoup"). Our analysis resulted in two main findings. First, we found that, when matched for lexical knowledge (vocabulary size) and lifetime oral language exposure (years of hearing experience), children with CIs and TH processed correctly-pronounced words along a similar timecourse. Differences between hearing groups instead stemmed from responses to mispronunciations: children with typical hearing tended to look equally to the familiar and unfamiliar objects (equivocating), or they looked more to the unfamiliar image (treating it as a novel word). In contrast, children with CIs preferred the familiar image. Thus, they showed reduced sensitivity to the mispronunciations and were more likely to disregard them. Second, for the children with CIs, sensitivity to mispronunciations was correlated with phonetic skill (articulatory accuracy on a standardized assessment), but not vocabulary size: children with higher articulation scores showed greater sensitivity to mispronunciations, in line with other work that has establish perception-production links in children of this age (Rvachew, n.d.; Shiller et al., n.d.).

Taken together, these results suggest that *all* children—those with and without CIs—use their lexical knowledge to process correct pronunciations. However, children rely on fine phonetic detail to process speech variation. In the absence of access to a rich, reliable phonetic signal, such as that generated from acoustic hearing, children do not develop the same sensitivity to speech variation. We elaborate upon these points below.

4.1 Articulatory skill, not vocabulary, predicts sensitivity

Contrary to our hypothesis, we did not find a reliable relationship between vocabulary size and mispronunciation sensitivity for either children with CIs or TH. This finding runs counter to previous work that has documented such a relationship in 2- to 3-year-olds (Law & Edwards, 2015; Swingley, 2016). In that work, vocabulary is cited as one possible mechanism that children may use to develop sensitivity to speech variation: children with larger vocabularies are thought to have more well-specified phonological representations due, in part, to the demands that denser phonological neighborhoods place upon representations (Edwards et al., 2004; Sosa & Stoel-Gammon, 2012; Stoel-Gammon, 2011).

Yet the modeling here did not demonstrate a relationship between children's vocabulary size and mispronunciation sensitivity—a relationship was only found for phonetic accuracy and then only for the children with CIs. For both children with CIs and TH, there was certainly sufficient variability between children to capture a potential effect of vocabulary (GSV score range 42-159 for children with CIs and 117-164 for children with TH, although all children with TH had above-average vocabulary sizes for their age). Consequently, differences between the current study and previous work could stem from the age group tested. Children in the current study are several years older than those previously studied, meaning that facilitative effects of vocabulary may only manifest within a certain developmental window. In further support of this idea are longitudinal data showing a facilitative effect of expressive vocabulary size for mispronunciation sensitivity at three years of age, but not four or five (Mahr, 2018). Thus, our null result of vocabulary is not at odds with previous work and is instead further evidence that vocabulary only predicts sensitivity to speech variation for a certain period in early development, before age 4.

Modeling did demonstrate a correlation between phonetic accuracy and mispronunciation sensitivity for the children with CIs. The children with higher articulation scores looked more quickly and reliably to the target word when they heard a correct pronunciation, acting quickly and decisively. This effect manifests visually in Figure 4 with the advantage of correct pronunciations over mispronunciations increasing with articulation ability. We interpret this finding as suggesting that the children who are skilled at capturing the phonetic signal during online processing—and are sensitive to disruptions in it—are the same children who are skilled at articulating sounds during speech production. At the age group studied (34–66 months), children who perform poorly on standardized tests of articulatory ability are no longer doing so *purely* for motoric reasons (i.e., inability to front the tongue dorsum). Instead, we believe that poor performance on both tasks (mispronunciation sensitivity and phonetic accuracy) indicates that a child with (a) CI(s) is less practiced at interpreting the electric hearing signal and manipulating it into a phonological representation that they can use in speech processing *and* production. The electric hearing that these children can access certainly allows them to learn and process words—after all, the children with CIs (as a group) processed correct pronunciations along a similar timescale as their vocabulary-matched peers with TH. Nevertheless, the fact that the standardized assessment of phonetic accuracy can explain variation in mispronunciation sensitivity suggests a single developmental mechanism underlying the children's ability to produce and perceive individual phonemes. The result suggests that well-specified phonological representations drive accuracy in speech production and sensitivity in speech processing.

4.2 Sensitivity to speech variation matters for word learning

As we have outlined in this paper, to learn words and phonemes, children must learn appropriate amounts of sensitivity to the speech used in their environments. Children must have access to a detailed (acoustic) speech signal to process variation. What happens when a degraded speech signal results in under-specified phonological representations? What are the consequences for learning to process speech variation, and thus learn new words? As we have shown here, vocabulary cannot help compensate for underspecified representations—at least after a certain stage in development. Even controlling for vocabulary size, children who classify novel phonological neighbors ("shoup") as variants on a word (*soup*) will struggle to map novel words to referents in their environments. These children might not consider variants like "shoup" to be completely homophonous with *soup*. The children with CIs studied here were, after all, sensitive to the mispronunciations. But perhaps they would require more exposure to a close phonological neighbor to map it to a referent in their environment. Perhaps these children would need to hear repeated exemplars of this close phonological neighbor, spoken by multiple interlocutors around them, in order to disentangle a potential variant of a known word from a new word to be mapped.

In either case, children would take longer to learn new words, especially those that fall into dense phonological neighborhoods. Unfortunately, the child lexicon is replete with dense phonological neighborhoods (cf. Charles-Luce and Luce (1990)). In typical development, children learn dense phonological neighborhoods *first* (Carlson, Sonderegger, & Bane, 2014; Jones & Brandt, 2019; Storkel, 2004), especially in production. So, for children systematically exposed to a degraded signal, one consequence could be the developmental trajectory of phonological neighborhood restructuring (Storkel, 2002). Charles-Luce and Luce (1990) originally postulated that children had sparser phonological neighborhoods than adults due to the underspecification of their phonological representations. In the decades since, research on sensitivity to mispronunciation among children with TH has shown that children *do* have relatively well-specified phonological representations (Swingley & Aslin, 2002; White & Morgan, 2008) and *do* learn dense neighborhoods (Storkel, 2004). However, for children with electric hearing, the current results suggest a developmental path more akin to that originally outlined in Charles-Luce and Luce (1990): electric hearing results in the kind of underspecified representations once proposed in Charles-Luce and Luce (1990) and thus children with CIs may have sparser neighborhoods than even their vocabulary-size matched peers with TH.

4.3 A note about the developmental trajectory of mispronunciation sensitivity

In the mispronunciation sensitivity paradigm, the most mature response to hearing a mispronunciation (“shoup”) is to look at the opposing image (i.e., to look *away* from an image of soup). This response indicates that a child noticed the /s/ > [ʃ] substitution and that the substitution disrupted their lexical access of *soup*.

However, as mentioned repeatedly in this work, sometimes efficient speech processing requires *ignoring* variation. Mature listeners and interlocutors regularly factor out speech variation stemming from differences in vocal tract morphology, speaking rate, and geographical dialect—the inability to do so would completely hinder communication. Now, a substitution such as /s/ > [ʃ] does, clearly, cross a phonemic threshold that phonetic variants on a word do not. For example, the frequency of /s/ lowers in rounded, back-vowel contexts, such as *soup*, and yet this variation does not disrupt word recognition. In fact, adult listeners know to compensate for the lowered /s/ in these environments (Mann & Repp, 1980). Still, while the *most* developmentally immature response to one phonological feature substitutions ("shoup") is to continue looking at the image of soup, suggesting that the child has not noticed the mispronunciation, and a *more* mature response is to recognize the mispronunciation and look away from the image of soup, the *most* mature response would be to (1) initially recognize the mispronunciation (look away from the soup) and then (2) recover from it (look back at the soup). This processing pattern would indicate that the child has recognized the mispronunciation—so they have relatively well-defined phonological representations—but has had sufficient experience processing speech to know to disregard some word variants. We call this an S-SHAPED RESPONSE to speech variation, a mature processing strategy where children would revise their original hypothesis concerning word identity. Mahr (2018), examining trials where the child started on the familiar image and heard a pronunciation, indeed found this mature strategy in typically developing 5-year-olds.

We were interested in exploring a potential s-shaped response in our own data. It seemed unlikely that such a pattern would emerge in the children with CIs who only have, on average, 33 months of hearing experience. So we instead explored the pattern in children with TH. These analyses were purely exploratory, not confirmatory, and future work should extend our analysis in a hypothesis-driven manner.

We again divided the children with TH into tertiles, in this instance by chronological age. Figure 6 plots raw response curves to the mispronunciation “shoup” with the timecourse of the audio stimulus. We plot this for the *soup*–“shoup” mispronunciation in particular because this substitution is phonetically grounded and observable in running speech: the frequencies of fricatives such as /s/ are known to lower (approximating [ʃ]) before back, round vowels. Our exploratory analysis shows that older children (58-66 months) exhibit more of the s-shaped response than either group of younger children, indicating that they revised their original lexical hypothesis. Younger children (36-48 and 49-57 months) have a flatter response with a slight increase in looking at the target image (soup) over time.

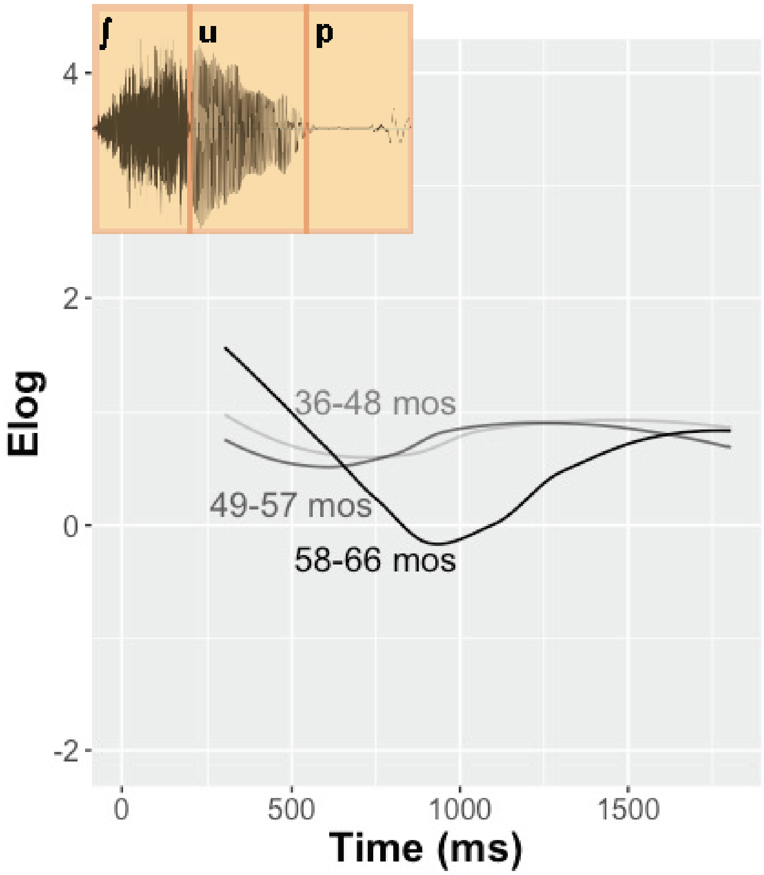


Figure 6. Raw response trajectories and audio stimulus for proportion of looks to soup upon hearing the mispronunciation "shoup" for children with TH. Children were divided into tertiles by chronological age.

These results are exploratory, but they do suggest that not only does overall sensitivity to mispronunciations increase with age (Von Holzen & Bergmann, 2021), but the processing strategy changes with age and cognitive maturity as well. Since a wealth of research has, by now, demonstrated that infants and children *are* sensitive to vowel and consonant mispronunciations, in various phonological environments, we now encourage future work examining the timecourse and processing strategies underling this sensitivity.

5 Conclusion

Variation in spoken language is rampant. Yet to learn the sounds and words used in the language spoken around them, young children must learn to contend with this variation. Using the unique sensory experiences of children who received cochlear implants (CIs), who hear via a degraded, electric speech signal, we demonstrated that children must have access to fine phonetic detail to process speech variation, starting in the early preschool years. Our results showed that 3- to 6-year-old children could rely on their lexical knowledge to process known words (*soup*), but experience with the degraded signal impacted how they processed variable pronunciations ("shoup"). Furthermore, children's ability to process variable pronunciations was correlated with their spoken phonetic accuracy, suggesting a single developmental mechanism underlying the ability produce *and* process individual phonemes. Thus, while children can rely on their lexical knowledge to process known words, they must have access to a robust speech signal, and well-specified phonological representations, to process variable pronunciations.

Acknowledgements

The authors thank the children and families who participated in this research study and acknowledge funding from the following sources: National Institute on Deafness and Other Communication Disorders grants F32DC019539 (M.C.) and R01DC02932 (B.M., J.E., & Mary E. Beckman). The authors additionally acknowledge the research assistants who facilitated data collection.

Author Contributions

MC, RN, BM, & JE designed the study. BM & JE collected the data. TM curated the data. MC analyzed the data and wrote the paper. All authors edited the paper.

Declarations of interest:

None.

Appendices

5.1 Appendix A

Table 6

Audiological information from the N=25 unique children with cochlear implants studied.

Participant	Matched to child with TH?	Chronological age	Age at hearing loss	Age at activation	Hearing age	Etiology	Device formation	Activation order
300E	Y	57	0	13	44	Genetic	Bilateral	simultaneous
302E	Y	37	0	13	24	Unknown	Bilateral	R-L
303E	Y	65	6	13	52	Unknown	Bilateral	simultaneous
304E	Y	48	0	12	36	Genetic	Bilateral	R-L
305E	Y	44	0	22	22	Unknown	Bilateral	R-L
306E	Y	49	0	8	41	Unknown	Bilateral	R-L
307E	Y	44	0	15	29	Genetic	Bilateral	R-L
309E	Y	59	0.5	7	52	Genetic	Bilateral	simultaneous
311E	Y	62	9	13	49	Unknown	Bilateral	L-R
314E	Y	38	10	17	21	Unknown	Bilateral	R-L
608L	Y	55	0.5	9	46	Connexin 26	Bilateral	simultaneous
665L	Y	40	0	12	28	Genetic	Bilateral	R-L
801E	Y	39	1.5	15	24	Unknown	Bilateral	simultaneous
804E	Y	56	0	7	49	Genetic	Bilateral	simultaneous
809E	Y	64	6	8	56	Meningitis	Bilateral	R-L
301E	N	53	0	45	8	Unknown	Bilateral	R-L
308E	N	37	0	13	24	Genetic	Bilateral	simultaneous
310E	N	52	unknown	23	29	Genetic	Bilateral	simultaneous
312E	N	57	0	24	33	Genetic	Bilateral	R-L
679L	N	58	0	29	29	Genetic	Bimodal	n/a
800E	N	65	30	37	28	Genetic	Bilateral	simultaneous
803E	N	41	0	34	7	Unknown	Bimodal	n/a
806E	N	42	14	34	8	Genetic	Unilateral	L
807E	N	51	10	22	29	Mondini malformation	Bimodal	n/a
808E	N	37	0	6	31	Genetic	Bilateral	simultaneous

6 References

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, *17*(2), 1265–1282. doi: 10.1016/S0885-2014(02)00116-8

Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language*, *32*(1), 159–173. doi: 10.1017/S0305000904006567

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. Retrieved 2022-03-22, from `https://linkinghub.elsevier.com/retrieve/pii/S0749596X07001015` doi: 10.1016/j.jml.2007.09.002

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. doi: 10.1016/j.cognition.2007.04.005

Bruggeman, L., Millasseau, J., Yuen, I., & Demuth, K. (2021). The Acquisition of Acoustic Cues to Onset and Coda Voicing Contrasts by Preschoolers With Hearing Loss. *Journal of Speech, Language, and Hearing Research*, *64*(12), 4631–4648. doi: 10.1044/2021_JSLHR-20-00311

Carlson, M. T., Sonderegger, M., & Bane, M. (2014). How children explore the phonological network in child-directed speech: A survival analysis of children's first word productions. *Journal of Memory and Language*, *75*, 159–180. doi: 10.1016/j.jml.2014.05.005

Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, *17*(1), 205–215. doi: 10.1017/S0305000900013180

Coretta, S. (2022). tidymv: Tidy model visualisation for generalised additive models [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=tidymv` (R package version 3.3.0)

Creel, S. C. (2012). Phonological similarity and mutual exclusivity: On-line recognition of atypical pronunciations in 3–5-year-olds. *Developmental Science*, *15*(5), 697–713. doi:

10.1111/j.1467-7687.2012.01173.x

Davidson, K., Lillo-Martin, D., & Chen Pichler, D. (2014). Spoken English Language Development Among Native Signing Children With Cochlear Implants. *Journal of Deaf Studies and Deaf Education*, *19*(2), 238–250. doi: 10.1093/deafed/ent045

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in novel word repetition. *Journal of Speech Language and Hearing Research*, *57*, 421–436.

Fagan, M. K. (2014). Frequency of vocalization before and after cochlear implantation: Dynamic effect of auditory feedback on infant behavior. *Journal of experimental child psychology*, *126*, 328–338. doi: 10.1016/j.jecp.2014.05.005

Fagan, M. K. (2015). Why repetition? Repetitive babbling, auditory feedback, and cochlear implantation. *Journal of experimental child psychology*, *137*, 125–136. doi: 10.1016/j.jecp.2015.04.005

Fenson, L., Marchman, V., Thal, D. J., Dale, P., Reznick, J., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories User's Guide and Technical Manual* (2nd Edition ed.). San Diego, CA: Singular.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking Up Speed in Understanding: Speech Processing Efficiency and Vocabulary Growth Across the 2nd Year. *Developmental psychology*, *42*(1), 98–116. doi: 10.1037/0012-1649.42.1.98

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernández, & H. Clahsen (Eds.), *Language Acquisition and Language Disorders* (Vol. 44, pp. 97–135). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/lald.44.06fer

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation-Second Edition (GFTA-2)* (Second ed.). Circle Pines, MN: American Guidance Service.

Houston, D. M., Stewart, J., Moberly, A., Hollich, G., & Miyamoto, R. T. (2012). Word learning in deaf children with cochlear implants: Effects of early auditory experience: Word learning in children with cochlear implants. *Developmental Science*, *15*(3), 448–461. doi: 10.1111/j.1467-7687.2012.01140.x

Huang, Y., & Snedeker, J. (2020). Evidence from the visual world paradigm raises questions about unaccusativity and growth curve analyses. *Cognition*, *200*, 104251. doi: 10.1016/j.cognition.2020.104251

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, *11*(6), F31-F39. doi: 10.1111/j.1467-7687.2008.00768.x

Jones, S. D., & Brandt, S. (2019). Do children really acquire dense neighbourhoods? *Journal of Child Language*, *46*(6), 1260–1273. doi: 10.1017/S0305000919000473

Law, F., & Edwards, J. R. (2015). Effects of Vocabulary Size on Online Lexical Processing by Preschoolers. *Language Learning and Development*, *11*(4), 331–355. doi: 10.1080/15475441.2014.961066

Mahr, T. (2018). *Development of word recognition in preschoolers* (Unpublished doctoral dissertation). University of Wisconsin, Madison, Madison, WI.

Mahr, T., & Edwards, J. (2018). Using language input and lexical processing to predict vocabulary size. *Developmental Science*, *21*(6), 1–14. doi: 10.1111/desc.12685

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, *57*(2), 252–272. doi: 10.1016/j.jml.2007.03.005

Mann, V., & Repp, B. (1980). Influence of vocalic context on perception of the [sh]-[s] distinction. *Perception & Psychophysics*, *28*, 213–228.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71–102. doi: 10.1016/0010-0277(87)90005-9

Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29–63.

Maslowski, M., Meyer, A. S., & Bosker, H. R. (2019). How the tracking of habitual rate influences speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 128–138. doi: 10.1037/xlm0000579

Newman, R. S., & Sawusch, J. R. (2009). Perceptual normalization for speaking rate III: Effects of the rate of one voice on perception of another. *Journal of Phonetics*, *37*(1), 46–65. doi: 10.1016/j.wocn.2008.09.001

Pisoni, D. B., Nusbaum, H. C., Luce, P. A., & Slowiaczek, L. M. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication*, *4*(1-3), 75–95. doi: 10.1016/0167-6393(85)90037-8

Pomper, R., Weismer, S. E., Saffran, J., & Edwards, J. (2019). Specificity of Phonological Representations for Children with Autism Spectrum Disorder. *Journal of autism and developmental disorders*, *49*(8), 3351–3363. doi: 10.1007/s10803-019-04054-5

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Reinisch, E. (2016). Speaker-specific processing and local context information: The case of speaking rate. *Applied Psycholinguistics*, *37*(6), 1397–1415. doi: 10.1017/S0142716415000612

Rescorla, L., & Ratner, N. B. (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-e). , *39*(1), 153–165. Retrieved from `http://pubs.asha.org/doi/10.1044/jshr.3901.153` doi: 10.1044/jshr.3901.153

Rij, J. V., Hollebr, B., & Hendriks, P. (2016). Children's eye gaze reveals their use of discourse context in object pronoun resolution. In *Empirical perspectives on anaphora resolution* (pp. 267–293).

Rvachew, S. (n.d.). Speech perception training can facilitate sound production learning. , *37*(2), 347–357. Retrieved 2022-05-04, from http://pubs.asha.org/doi/10.1044/jshr.3702.347 doi: 10.1044/jshr.3702.347

Sekhon, J. S. (2011). Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The {Matching} Package for {R}. *Journal of Statistical Software*, *42*(7), 1–52.

Shiller, D. M., Rvachew, S., & Brosseau-Lapré, F. (n.d.). Importance of the auditory perceptual target to the achievement of speech production accuracy. , *34*(3), 181–192.

Sosa, A. V., & Stoel-Gammon, C. (2012). Lexical and Phonological Effects in Early Word Production. *Journal of Speech Language and Hearing Research*, *55*(2), 596. doi: 10.1044/1092-4388(2011/10-0113)

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, *84*, 101017. doi: 10.1016/j.wocn.2020.101017

Steffman, J. (2019). Intonational structure mediates speech rate normalization in the perception of segmental categories. *Journal of Phonetics*, *74*, 114–129. doi: 10.1016/j.wocn.2019.03.002

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children*. *Journal of Child Language*, *38*(01), 1–34. doi: 10.1017/S0305000910000425

Storkel, H. L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language*, *29*(2), 251–274. doi: 10.1017/S0305000902005032

Storkel, H. L. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, *25*, 201–221. doi: 10.1017/S0142716404001109

Swingley, D. (2016). Two-year-olds interpret novel phonological neighbors as familiar words. *Developmental Psychology*, *52*(7), 1011–1023. doi: 10.1037/dev0000114

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147–166. doi: 10.1016/S0010-0277(00)00081-0

Swingley, D., & Aslin, R. N. (2002). Lexical Neighborhoods and the Word-Form Representations of 14-Month-Olds. *Psychological Science*, *13*(5), 480–484. doi: 10.1111/1467-9280.00485

Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., & Demuth, K. (2019). The Acquisition of Mandarin Tonal Processes by Children With Cochlear Implants. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1309–1325. doi: 10.1044/2018_JSLHR-S-18-0304

Tao, Y., Xu, Q.-M., & Plunkett, K. (2012). Phonological specificity of tones in early lexical representation at 16 months of age. *Acta Psychologica Sinica*, *44*(8), 1066–1074.

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2020). *itsadug: Interpreting time series and autocorrelated data using gamms.* (R package version 2.4)

Vihman, M. M. (2014). *Phonological development: The first two years* (2nd ed.). Wiley-Blackwell.

Von Holzen, K., & Bergmann, C. (2021). The development of infants' responses to mispronunciations: A meta-analysis. *Developmental Psychology*, *57*(1), 1–18. doi: 10.1037/dev0001141

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152. doi: 10.1177/0956797613488145

Wewalaarachchi, T. D., & Singh, L. (2020). Vowel, consonant, and tone variation exert asymmetrical effects on spoken word recognition: Evidence from 6-year-old monolingual and bilingual learners of Mandarin. *Journal of Experimental Child Psychology*, *189*, 104698. doi: 10.1016/j.jecp.2019.104698

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations.

Journal of Memory and Language, *59*(1), 114–132. doi: 10.1016/j.jml.2008.03.001

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. Retrieved from `https://ggplot2.tidyverse.org`

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116. doi: 10.1016/j.wocn.2018.03.002

Wilke, C. O. (2020). cowplot: Streamlined plot theme and plot annotations for 'ggplot2' [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=cowplot` (R package version 1.1.1)

Williams, K. (2007). *Expressive Vocabulary Test, Second Edition* (Second ed.). San Antonio, TX: Pearson Education.

Winn, M. B., & Litovsky, R. Y. (2015). Using speech sounds to test functional spectral resolution in listeners with cochlear implants. *The Journal of the Acoustical Society of America*, *137*(3), 1430–1442. doi: 10.1121/1.4908308

Wood, S. (2017a). *Generalized additive models: An introduction with R* (2nd Edition ed.). Boca Raton: CRC Press.

Wood, S. (2017b). *Generalized additive models: An introduction with r* (2nd ed.). Chapman and Hall/CRC.

Wood, S., Pya, N., & Saefken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, *111*, 1548–1575.

Zahner, K., Kutscheid, S., & Braun, B. (2019). Alignment of f0 peak in different pitch accent types affects perception of metrical stress | Elsevier Enhanced Reader. *Journal of Phonetics*, *74*, 75–95. doi: 10.1016/j.wocn.2019.02.004