ABSTRACT

Title of Dissertation:	SPECTRAL CONTRASTS PRODUCED BY CHILDREN WITH COCHLEAR IMPLANTS: INVESTIGATING THE IMPACT OF SIGNAL DEGRADATION ON SPEECH ACQUISITION
	Allison Ann Johnson, Doctor of Philosophy, 2022
Dissertation directed by:	Professor Jan R. Edwards, Department of Hearing and Speech Sciences

The primary objective of this dissertation was to assess four consonants, /t/, /k/, /s/, and /ʃ/, produced by young children with cochlear implants (CIs). These consonants were chosen because they comprise two place-of-articulation contrasts, which are cued auditorily by spectral information in English, and they cover both early-acquired (/t/, /k/) and late-acquired (/s/, /ʃ/) manners of articulation. Thus, the auditory-perceptual limitations imposed by CIs is likely to impact acquisition of these sounds: because spectral information is particularly distorted, children have limited access to the cues that differentiate these sounds.

Twenty-eight children with CIs and a group of peers with normal hearing (NH) who were matched in terms of age, sex, and maternal education levels participated in this project. The experiment required children to repeat familiar words with initial /t/, /k/, /s/, or /ʃ/ following an auditory model and picture prompt. To

create in-depth speech profiles and examine variability both within and across children, target consonants were elicited many times in front-vowel and back-vowel contexts. Patterns of accuracy and errors were analyzed based on transcriptions. Acoustic robustness of contrast was analyzed based on correct productions. Centroid frequencies were calculated from the release-burst spectra for /t/ and /k/ and the fricative noise spectra for /s/ and /ʃ/.

Results showed that children with CIs demonstrated patterns not observed in children with NH. Findings provide evidence that for children with CIs, speech acquisition is not simply delayed due to a period of auditory deprivation prior to implantation. Idiosyncratic patterns in speech production are explained in-part by the limitations of CI's speech-processing algorithms.

The first chapter of this dissertation provides a general introduction. The second chapter includes a validation study for a measure to differentiate /t/ and /k/ in adults' productions. The third chapter analyzes accuracy, errors, and spectral features of /t/ and /k/ across groups of children with and without CIs. The fourth chapter analyzes /s/ and /ʃ/ across groups of children, as well as the spectral robustness of both the /t/-/k/ and the /s/-/ʃ/ contrasts across adults and children. The final chapter discusses future directions for research and clinical applications for speech-language pathologists.

SPECTRAL CONTRASTS PRODUCED BY CHILDREN WITH COCHLEAR IMPLANTS: INVESTIGATING THE IMPACT OF SIGNAL DEGRADATION ON SPEECH ACQUISITION

by

Allison Ann Johnson

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
Dr. Jan Edwards, Chair
Dr. Rochelle Newman
Dr. Matthew Goupell
Dr. Margaritis Fourakis
Dr. Stefanie Kuchinsky
Dr. Catherine Carr, Dean's Representative

Dedication

This dissertation is dedicated to all the parents who put their trust in science when faced with the impossibly difficult decision about whether to permanently implant an electronic device into their infant's skull. This research represents a small part of the enormous progress that would not have been possible without your confidence and investment in speech scientists like me.

Acknowledgements

Although this dissertation is considered a presentation "my own research," I could not have completed any aspect of this project by myself.

I have the deepest, most sincere gratitude for my mentor (and hero), Dr. Jan Edwards. Even though she uprooted my life by moving us across the country halfway through my program, she is still, without a shadow of a doubt, the BEST advisor on Earth. If I weren't writing this at the last possible minute, I could easily fill 100 more pages describing all the ways that Jan has supported me since I joined the Learning to Talk lab as an undergraduate. She has influenced my personal and intellectual development more over the past 10 years than my own mother. Thank you for investing in me and providing the aggressive encouragement that I needed to get through this most challenging endeavor of my life.

Thanks to the members of my dissertation committee, Dr. Catherine Carr, Dr. Marios Fourakis, Dr. Matt Goupell, Dr. Stefanie Kuchinsky, and Dr. Rochelle Newman, as well as Dr. Mary Beckman, Dr. Ben Munson, Dr. Pat Reidy, Dr. Gary Weismer, and Dr. Nan Ratner for their input, feedback, and guidance at various stages of this project. Thanks also to the professors who inspired me to pursue a career in science and devoted an enormous amount of time imparting their knowledge and sharing their resources: Dr. Rita Kaushanskaya, Dr. John Westbury, Dr. Susan Ellis Weismer, Dr. Jenny Saffran, Dr. Nadine Connor, Dr. Michelle Ciucci, Dr. Katie Hustad, and Dr. Audra Sterling.

Shoutout (SO) to the entire Learning to Talk Lab, both current and former members, as well as all of the children and families who participated in our lab's work. Thanks especially to Tatiana Thonesavanh, Nancy Wermuth, Dr. Tristan Mahr, Dr. Matt Winn, Dr. Franzo Law II, Megan Flood, Kayla Kristensen, Michelle Minter, Alissa Schneeberg, Ruby Braxton, Nicole Rohena, Michelle Erskine, Danielle Bentley, Sara Cline, Becky Johnson, Bianca Behnke, Lizzy Hill, Hyuana Kim, Emily Ganser, and the other experimenters who helped collect or code data.

Thanks to my UMD compatriots, Zach Maher, Christina Blomquist, Arynn Byrd, Kathleen Oppenheimer, Dr. Meg Cychosz, Dr. Julianne Garbarino, Dr. Anna Tinnemore, Dr. Mary Barrett, and the entire "Nailed it!" trivia team for helping me build a new life in Maryland. Also, to Dr. Colin Phillips, Dr. Shevaun Lewis, Dr. Tess Wood, Caitlin Eaves, and all the members of the Language Science Center who taught me that "Science is Social."

Thanks to my medical team, for helping me recover from a severe concussion while finishing this dissertation: Karen Hennessey, Amanda Mendizaball, Kristin Clemens, Dr. Trey Godwin, Dr. Nicole Fromm, Dr. Aubrey Verdun, Kelly Masterson, Betty Chan, and Dr. Michael Kotlicky.

Thanks to my sister and my friends, who contacted me every single day over the past year to make sure I was ok and motivate me to persevere during the most

stressful times. Thanks to my parents, my grandparents, and my in-laws for their unwavering support and encouragement.

Finally, thanks to my adoring husband, Jake Cox. I could write another 100+ pages detailing all the ways that you have helped me get through this process, such as quitting your job and moving away from your entire family and all of your friends to come live with me in Maryland, but I think we'd both prefer if I just stop writing now so we can watch the Bucks together.

This research was funded by: NIDCD R01 02932 grant to Jan Edwards, Mary E. Beckman, and Benjamin Munson; NICHD P30 HD03352 grant to the Waisman Center; T32 Training Grant DC 05359-10 to Susan Ellis Weismer; T32 Training Grant DC-00046 to Catherine Carr & Sandra Gordon-Salant; and NSF grant #1449815 to Colin Phillips.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Children born with profound hearing loss (approximately 2 infants per 1,000 born) who are fitted with cochlear implants (CIs) typically have the long-term intervention goals of communicating orally and participating alongside their peers with normal hearing (NH) in mainstream education and activities (Young & Kirk, 2016). Although CIs are arguably the most sophisticated medical prostheses available, they have speech processing limitations and do not precisely mimic the auditory processing of a healthy cochlea. CIs do not restore "normal" hearing (for review, see Loizou, 2006; Wilson, 2004; Zeng, 2008). As a result, young children using CIs face unique challenges while learning to listen, perceive, and produce speech sounds accurately. Even when children participate regularly in specialized communication training after receiving their implants, they continue to demonstrate delays in speech development compared to peers with NH (e.g., Blamey, Barry, & Jacq, 2001; Spencer & Guo, 2012).

The purpose of this dissertation is to examine the acquisition of two consonant contrasts (/t/ vs. /k/ and /s/ vs. /ʃ/) across groups of 3- to 5-year-old, English-learning children with and without cochlear implants learning English in order to quantify the consequences of signal degradation on speech acquisition, specifically speech-sound production. This research will not only guide engineers working to improve speech processing algorithms, but also inform speech-language pathologists designing more focused interventions to facilitate speech acquisition despite unique perceptual constraints.

The two voiceless consonant contrasts of interest in this study were assessed using both transcription and acoustic measures. These sounds were selected for several reasons. First, successful production is likely affected by both *universal* developmental processes (e.g., motor control) as well as *unique* signal degradation processes (e.g. frequency distortions). The sibilant fricatives (/s/ and /ʃ/) are typically acquired later than the stops (/t/ and /k/) because they require more advanced neuro-motor control. For 3- to 5-year-old children, productions of /s/ and /ʃ/ are likely to be highly variable—both within and across children—due to ongoing maturation. Because high variability is more likely to obscure group differences for the late-acquired fricatives, comparing performance on an early-acquired contrast will help demarcate the unique effects of signal degradation. Second, these consonants comprise two place-of-articulation contrasts. Place contrasts are differentiated auditorily by spectral cues, and these specific sounds contain energy in the high-frequency range. When sound is processed by CIs, spectral information is particularly degraded, and energy above 8,000Hz is eliminated entirely; thus, the sounds in these contrasts may be particularly challenging for children with CIs to differentiate. Third, these consonants involve articulatory gestures that are not straightforward to perceive visually: the tongue forms a constriction near the alveolar ridge (for /t/ and /s/) or farther back in the mouth toward the hard and soft palates (for /k/ and /ʃ/). Compared to sounds that involve the lips and/or teeth, children learning these sounds must rely on auditory information more than visual information. Thus, production patterns can be attributed more easily to auditory limitations. Finally, these sounds are frequent in the words of young children's vocabularies, so performance should not be affected by

a lack of exposure or too few opportunities to hear and practice the sounds in a variety of contexts.

The remaining chapters of this dissertation describe both published (/t/ vs. /k/) and unpublished (/s/ vs. /ʃ/) research conducted as part of this large, multi-faceted project. Chapter 2 presents a proof-of-concept study, where adults completed the same experimental task in order to validate a method of quantifying robustness of the /t/-/k/ contrast using a single, static spectral measure that could also be applied to the /s/-/ʃ/ contrast. Chapters 3 and 4 apply this validated measure to children's productions, and between-group comparisons are made in terms of overall accuracy, patterns of errors, and robustness of contrast. Chapter 4 also includes analyses of robustness of both the /t/-/k/ and the /s/-/ʃ/ contrasts across adults and children with and without CIs. The final chapter includes a general discussion and conclusions.

Chapter 2: Quantifying robustness of the /t/-/k/ contrast using a single, static spectral feature[1]

Abstract

Dynamic spectral shape features accurately classify /t/ and /k/ productions across speakers and contexts. This paper shows that word-initial /t/ and /k/ tokens produced by 21 adults can be differentiated using a single, static spectral feature when spectral energy concentration is considered relative to expectations within a given speaker and vowel context. Centroid and Peak frequency—calculated from both acoustic and psychoacoustic spectra—were compared to determine whether one feature could reliably differentiate /t/ and /k/, and, if so, which feature best differentiated them. Centroid frequency from both acoustic and psychoacoustic spectra accurately classified productions of /t/ and /k/.

Introduction

Over the past several decades, researchers have focused on classifying stop consonants in order to identify invariant, acoustic cues to place of articulation. This work was initially driven by a theoretical interest in how listeners achieve perceptual constancy despite variability in the acoustic signal (e.g., Blumstein & Stevens, 1979; Kewley-Port, 1983; Kewley-Port & Luce, 1984; Stevens & Blumstein, 1978.) The search for invariant features was further inspired by a practical interest in improving automatic speech recognition systems (e.g., Nossair & Zahorian, 1991). Clinical

[1] Reproduced from Johnson, A.A., Reidy, P.F., & Edwards, J. (2018). Quantifying robustness of the /t/-/k/ contrast using a single, static spectral feature. *Journal of the Acoustical Society of America, 144*(2), E105-E111, with the permission of AIP Publishing)

researchers have also expressed the need for a classification approach to quantify the spectral distance between model token productions and atypical ones (e.g., Forrest et al., 1988). Quantifying an individual speaker's robustness of contrast would support more reliable measures of covert contrasts, improve tracking of developmental changes, and provide more fine-grained descriptions of productions (compared to categorical transcription) upon which to compare groups. Our ultimate goal is to establish a clinically viable method for quantifying the degree of overlap between two sound categories within a speaker. As a first step toward this goal, we focus here on fluent productions of target /t/ and /k/ by healthy adult speakers, which should be differentiable using some feature(s) computed from the spectra of these productions. The purpose of this paper is to determine whether a single, static spectral feature, such as centroid or peak frequency, can sufficiently classify /t/ and /k/ tokens, given knowledge of the speaker and vowel context, and to compare features computed from acoustic versus psychoacoustic spectra.

There is a general consensus in the literature that critical information for differentiating place of articulation in English stop consonants is concentrated in the release burst. Early work by Stevens and Blumstein (1978; Blumstein & Stevens, 1979) identified a unique spectral shape of the burst for each place of articulation, including diffuse-rising for alveolars and a compact, mid-frequency spectral peak for velars. Kewley-Port (1983) expanded on these "templates" with time-varying features, and Kewley-Port and Luce (1984) further improved classification accuracy by demarcating speaker-specific values for "mid-frequency." Limitations of this early work include few speakers, poor generalizability across speakers, and less reliable

classification across different vowel contexts. Furthermore, these early classification schemes relied on time-consuming human visual judgments rather than objective methods.

Time-varying spectral features have since been quantified and gained additional support in the automatic classification literature. Nossair and Zahorian (1991) identified stop consonants with 93.7% accuracy from 20 dynamic, global features (discrete cosine transform coefficients) extracted from a 60-ms window around the stop burst. They achieved this high classification accuracy for 30 different speakers (including men, women, and children), across both voiced and voiceless consonants in a variety of vowel contexts.

Forrest et al. (1988) also achieved a high classification accuracy (93%) for voiceless stops across 10 speakers using only three spectral features (mean, skew, and kurtosis) extracted from a series of analysis windows spanning 40 ms. Relatively high spectral kurtosis was a defining feature for /k/, and negative spectral skew was the essential feature for /t/. The inclusion of skew and kurtosis possibly improved classification accuracy across speakers because these moments remove differences in spectral means that arise between speakers producing the same target sound, which accomplishes a rough speaker normalization (see Forrest et al., 1988, p. 118).

For Nossair and Zahorian's (1991) purpose of developing a speaker-independent, automatic speech classifier, a high-dimensional feature space is well motivated. However, a low-dimensional feature space—or even a single feature—may be more feasible for clinicians to obtain and interpret. A common theme among previous works is that multiple features calculated over relatively long analysis

windows have been necessary to achieve accurate and reliable classification of stop consonants across speakers and vowel contexts. However, speaker characteristics and coarticulation influence spectral shapes (and the features computed therefrom). The location of "mid-frequency" varies depending on an individual's vocal tract. Similarly, spectral peak frequency and spectral kurtosis can fluctuate within a speaker for /k/ in front- versus back-vowel contexts. The spectrum for /k/ can present with two spectral peaks due to resonances in both the front and back oral cavities. The spectrum for /t/ can also have a prominent peak near the speaker's F2 locus, and the energy concentration can shift if the tongue dorsum raises in preparation for a high front vowel.

Regardless of vowel context or speaker, /k/ is formed farther back in the mouth than /t/. Thus, theoretically, the overall concentration of energy in the spectrum for /k/ should be lower than that for /t/ for a particular speaker in a given vowel context. McMurray and Jongman (2011) recently conducted a comprehensive review of acoustic features for fricatives and found that none were entirely invariant. They identified fricatives using a compensation model (C-CuRE: Computing Cues Relative to Expectations), which used hierarchical regressions to capitalize on acoustic variability in the signal, and adjusted category expectations relative to known indexical and contextual information (such as speaker identity and vowel context).

Given the current power of mixed-effects modeling, we now have the capacity to process large sets of non-independent observations and examine variability both within and across speakers (for an overview on mixed-effects modeling, see Brauer & Curtin, 2017). It's possible that when speaker and vowel context are statistically

controlled, a single static spectral feature calculated over a relatively short analysis window will be sufficient to classify /t/ and /k/ tokens.

Forrest et al., (1988) successfully used centroid frequency in combination with skew and kurtosis to classify stop consonants. Because energy concentration should be at higher frequencies for /t/ than /k/ due to different places of articulation, centroid frequency may be suitable to differentiate /t/ from /k/ within a speaker. On the other hand, centroid frequency may not characterize the frequency-location of energy concentration in the spectrum very well if the distribution is bimodal. Thus, the frequency of the most prominent peak (henceforth, "peak frequency") may provide better evidence for the location of the constriction.

With the goal of identifying psychoacoustically relevant features, some researchers have also explored the effect of transforming an acoustic spectrum prior to computing features from it. Typically, these transformations seek to model some process of the auditory system, such as compression of the frequency scale or wider bandwidths at higher frequencies. For example, Forrest et al. (1988) transformed the hertz frequency scale of acoustic spectra to the Bark frequency scale, but did not apply any transformation to model the different bandwidths of auditory filters; classification of voiceless stops was poorer when features were computed from Bark-scale spectra than from hertz-scale spectra. Kewley-Port and Luce (1984) passed acoustic spectra through a bank of bandpass filters that modeled the different bandwidths of auditory filters, then transformed the hertz scale to the mel scale; however, classification accuracy from transformed spectra was not compared to

untransformed spectra, so it is difficult to assess the utility of these transformations on the classification of voiceless stops.

We address this lacuna by computing two features—Centroid and Peak frequency—from both acoustic spectra and from transformed spectra (henceforth, "psychoacoustic spectra") that were passed through a gammatone filter bank that models both the frequency-scale compression and the differential frequency selectivity of the auditory system. Our purpose is to determine whether a single, static feature can sufficiently differentiate /t/ and /k/ productions when speaker identity and vowel context are statistically controlled. This research is driven by the need for a standardized approach to quantify robustness of an individual's /t/-/k/ contrast that can be applied quickly, easily, and objectively by researchers and clinicians alike.

Methods

Twenty-one adult participants (10 women, 11 men; mean age: 21 years, range: 20-29 years) were recruited to participate from Minneapolis, MN. All participants were monolingual, native speakers of Mainstream American English with self-reported normal hearing and no history of speech or language disorders.

Stimuli for the experiment—a picture-prompted, auditory word repetition task—consisted of familiar words presented in isolation. Stimuli were recorded by an adult female speaker in a sound-treated lab setting, and recordings were normalized for amplitude. Words were also represented visually by archetypal, high-quality photographs obtained from online sources and edited for consistency in size and background. Stimuli were organized into two wordlists, each with 16 /t/-initial and 16

/k/-initial tokens. Each wordlist also included either 58 or 94 filler-words that did not begin with /t/ or /k/. Tokens were balanced across front- and back-vowel contexts. Possible front vowels included /i ɪ e ɛ æ/. Possible back vowels included /u ʊ o ʌ ɑ/ and the diphthongs /aɪ aʊ/. Despite known regional variations, our speakers and participants consistently produced these diphthongs with an initial back vowel.

All testing was completed in a sound-treated recording booth. During the experimental task, participants sat in front of a computer screen positioned approximately six inches away from a Shure SM81 cardioid condenser microphone with a custom pop filter. Words were presented in a pseudo-randomized order across participants, with steps taken to ensure target words were not repeated on consecutive trials. Visual stimuli appeared on the screen while auditory stimuli played over loudspeakers. At word-offset, participants repeated the word into the microphone, and an experimenter recorded the session using a Marantz PMD671 solid-state recorder at a sampling frequency of 44,100 Hz.

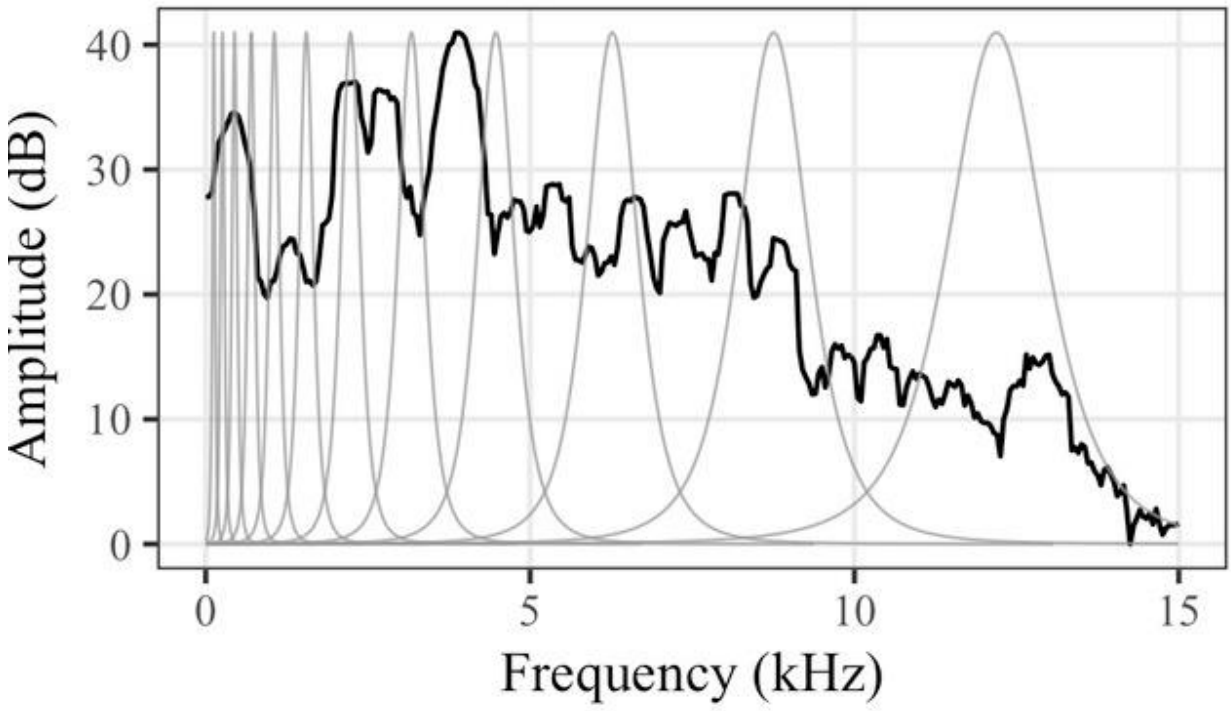
Five participants completed one wordlist (32 productions per speaker), and sixteen participants completed both wordlists (64 productions per speaker). Productions were excluded from analyses if there was background noise obscuring the release burst, or if voice-onset time was less than 20-ms. The final number of analyzable tokens was 1155.

Coding was done in *Praat* (Boersma & Weenink, 2018). The first author transcribed place of articulation for all /t/ and /k/ tokens. Then, she marked locations on the waveform corresponding to the release burst and the onset of voicing. The release-burst was defined as the first transient-noise spike following a period of

silence that coincided perceptually with the release of an oral constriction. The onset of voicing was defined as the first upward swing from the zero-crossing followed by a stable, quasi-periodic pattern of voicing. A second trained phonetician coded a random 20% of the files for reliability purposes. Reliability between the two coders was high: agreement for transcriptions was 100%. Root-mean-square (RMS) values were calculated to determine differences in locations of burst and VOT tags. For burst locations, RMS error was 0.0023 ms, and for VOT locations, RMS error was 0.0039 ms.

Acoustic and statistical analyses were carried out in the *R* programming environment (R Core Team, 2013), using custom scripts. The method for computing acoustic and psychoacoustic spectra was identical to that reported in Reidy (2016), to which the reader is referred for a comprehensive description with references. For each token, 5 ms prior to the burst tag through 20 ms after the tag defined a 25-ms analysis window. Within this window, the acoustic spectrum of the waveform was estimated with an 8th-order multitaper spectrum. To transform an acoustic spectrum into a psychoacoustic spectrum, it was passed through a filter bank that modeled how the auditory periphery logarithmically compresses the frequency scale and how it differentially resolves frequency components across the audible range (see top panel of Fig. 1). This filter bank comprised 361 fourth-order gammatone filters whose center frequencies were equally spaced every 0.1, from 3 to 39, along the ERB_N number scale (Glasberg & Moore, 1990). The bandwidth of each filter was set to 1.019 times the equivalent rectangular bandwidth of that filter's center frequency in hertz. Each gammatone filter acted on an input spectrum as a bandpass filter. Finally,

the psychoacoustic spectrum was constructed by summing the total energy (or "auditory excitation") at the output of each filter and plotting these excitation levels against the filters' center frequencies in ERB_N. (see bottom panel of Fig. 1).



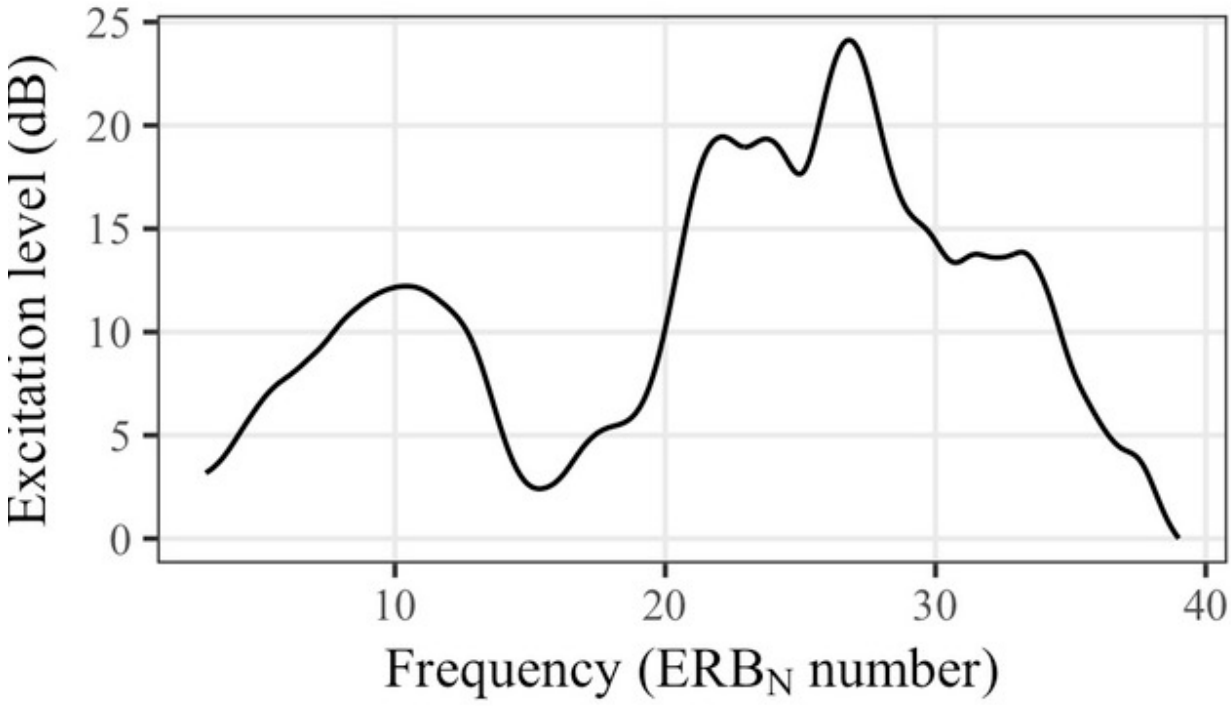


Figure 1: (Top) Acoustic spectrum of a production of /t/ estimated with an 8th-order multitaper spectrum (Centroid = 3.740 kHz, Peak = 3.918 kHz). Frequency responses of 12 gammatone filters of different center frequencies are shown, in grey, overlaid on the spectrum. (Bottom) The psychoacoustic spectrum resulting from passing the spectrum in the top panel through a 361-channel gammatone filter bank model of the auditory periphery (Centroid = 26.0 ERB_N, Peak = 26.8 ERB_N).

Centroid and Peak frequency were computed from both the acoustic spectra (within the .926-9.777 kHz range) and the psychoacoustic spectra (within the 15-35 ERB_N number range). To compute Centroid frequency, the values of a (psycho)acoustic spectrum were normalized so they summed to 1. The normalized (psycho)acoustic spectrum was then treated as a probability mass function over frequency, and the Centroid frequency was the distribution's mean value. Peak frequency was the frequency of the (psycho)acoustic spectrum with the greatest amplitude. Thus, there were four features computed from each token: Centroid (kHz), Centroid (ERB_N), Peak (kHz), and Peak (ERB_N). Prior to statistical analysis, the

values for each feature were centered by subtracting the group mean value for that feature.

Our modeling procedure followed that of Holliday et al. (2015), which quantifies the degree of category overlap within an individual speaker. We used four mixed-effects logistic regression models—one for each spectral feature—to predict each token's Target Consonant (either /t/ or /k/). Then, we used two additional mixed-effects logistic regression models to formally compare the accuracy of predictions made by each model. All models were fit using the *R* lme4 package (Bates, Mächler, Bolker, & Walker, 2014, v. 1.1-11).

Logistic regression models are based on the logarithm function, and they are an appropriate statistical choice when the outcome variable is binary or binomially distributed, as in this case where the outcome variable is a prediction of either /t/ or /k/ (for more on analyzing categorical data, see Jaeger, 2008). The dependent variable in a logistic regression model is a log-likelihood ratio, which can be used to determine the probability of one outcome or the other given values of the independent variables. Mixed-effects models, which refer to models that contain at least one conditional random effect, are appropriate to use when data are non-independent, as in this case when multiple tokens are obtained from each speaker. Random effects by-participant produce participant-level adjustments for predictors, which can be used to obtain unique, individually-fit models for each participant (for more information on mixed-effects modeling, see Bates et al., 2015, or Brauer & Curtain, 2017).

An example of the formula used to make predictions is shown in Eq. (1)[2], where the subscripts *i* and *j* range over items and speakers, respectively.

$$log\left(\frac{/t/}{1-/t/}\right)_{ij} = \beta_0 + \beta_1 \times Centroid(kHz)_{ij} + \beta_2 \times VowelContext_{ij} + \beta_3 \times Centroid(kHz)_{ij} \times VowelContext_{ij} + u_{0j} + u_{1j} \times Centroid(kHz)_{ij} + \varepsilon_{ij} \quad (1)$$

This model predicted the log-likelihood that the target consonant for a given token was /t/ based on fixed effects of the group-wide intercept (β_0), Centroid frequency computed from acoustic spectra (β_1), Vowel Context (β_2), the interaction between Centroid frequency and Vowel Context (β_3), as well as the speaker-level random intercept (u_{0j}) and slope for Centroid frequency (u_{1j}). Three additional models with homologous structures to that in Eq. (1) were fit to make predictions based on the other spectral features of interest. When the predicted log-likelihood was greater than 0, the model predicted the target consonant to be /t/; otherwise, the model classified the token as /k/. If the prediction matched the target consonant, the token was assigned a 1 for *Predicted Accuracy*. If the model made an incorrect prediction, *Predicted Accuracy* was 0. Predictions made by each of the four models were highly accurate. Results are shown in Table 1.

[2] *R* code for implementing Eq. (1): `glmer(formula = Target Consonant ~ Centroid.kHz * Vowel Context + (1 + Centroid.kHz | Participant), data = Adult Productions, family = 'binomial')`

Table 1. Overall accuracy of predictions made by each model (one model for each spectral feature), and the accuracy of predictions by target consonant and vowel context.

Spectral Feature	**Target /t/**		**Target /k/**	
Centroid (kHz)	**Front**	**Back**	**Front**	**Back**
95%	94%	95%	95%	94%
Centroid (ERB_N)	**Front**	**Back**	**Front**	**Back**
95%	91%	98%	94%	96%
Peak (kHz)	**Front**	**Back**	**Front**	**Back**
89%	83%	87%	93%	93%
Peak (ERB_N)	**Front**	**Back**	**Front**	**Back**
90%	84%	94%	91%	90%

To determine which of the four spectral features best differentiated /t/ and /k/, we ran two additional mixed-effects logistic regression models that compared accuracy of predictions. We added two variables to our dataset: *Representation* and *Feature*. *Representation* was either 'Acoustic' or 'Psychoacoustic,' referring respectively to whether the spectral feature was computed from an acoustic or psychoacoustic spectrum. *Feature* was either 'Centroid' or 'Peak.' The formula comparing accuracy of predictions for the spectral features is shown in Eq. (2)[3].

[3] *R* code for implementing Eq. (2): `glmer(formula = Predicted Accuracy ~ Representation * Feature + (1 | Participant), data = Adult Productions, family = 'binomial')`

$$log(\frac{PredictedAccuracy}{1-PredictedAccuracy})_{ij} = \beta_0 + \beta_1 \times Representation_{ij} + \beta_2 \times Feature_{ij} + \beta_3 \times Representation_{ij} \times Feature_{ij} + u_{0j} + \varepsilon_{ij}$$

(2)

Model 1 and Model 2 predicted the log-likelihood that a prediction was accurate based on fixed effects of group-wide intercept (β_0), Representation (β_1), Feature (β_2), the interaction between Representation and Feature (β_3), and the speaker-level random intercept (u_{0j}).

The difference between the two models was the reference category. In Model 1, the reference level for Representation was 'Psychoacoustic,' and the reference level for Feature was 'Centroid.' For Model 1 with Centroid (ERB_N) as the reference category, the main effect of Representation characterized the difference in accuracy of predictions based on Centroid (ERB_N) versus Centroid (kHz), and the main effect of Feature characterized the difference in predictions based on Centroid (ERB_N) versus Peak (ERB_N). In Model 2, the reference category was Peak (kHz), so the main effect of Representation characterized the difference between Peak (kHz) and Peak (ERB_N), and the main effect of Feature characterized the difference between Peak (kHz) and Centroid (kHz). Because we ran two, re-leveled models testing the same data, we used an adjusted alpha-level, $p = 0.0025$, to denote significance.

Results[4]

Model 1 (reference: Centroid (ERB_N)) showed significant main effects of intercept ($\hat{\beta}_0$ = 3.44, *SE* = 0.25, *z* = 13.72, *p* < 0.001) and Feature ($\hat{\beta}_2$ = -0.75, *SE* = 0.17, *z* = -4.52, *p* < 0.001). The main effect of Representation and the interaction were not significant. The results of Model 1 indicate that accuracy of predictions decreased significantly when Peak (ERB_N) was used compared to Centroid (ERB_N), but there was no difference between Centroid (ERB_N) and Centroid (kHz).

Model 2 (reference: Peak (kHz)) showed significant main effects of intercept ($\hat{\beta}_0$ = 2.58, *SE* = 0.23, *z* = 11.13, *p* < 0.001) and Feature ($\hat{\beta}_2$ = 0.83, *SE* = 0.16, *z* = 5.09, *p* < 0.001). The main effect of Representation and the interaction were not significant. The results of Model 2 indicate that accuracy of predictions increased significantly when Centroid (kHz) was used compared to Peak (kHz), but there was no difference between Peak (kHz) and Peak (ERB_N).

Taken together, these results suggest that Centroid frequency better differentiated /t/ and /k/ than Peak frequency, but there was no difference between spectral features computed from acoustic spectra versus psychoacoustic spectra.

Discussion

The central finding of this paper is that word-initial /t/ and /k/ tokens in the context of 12 different vowels produced by 21 different speakers were differentiated

[4] Detailed model results for all analyses included in this chapter are included in Appendix A; original text retained.

with 95% accuracy using a single, static, spectral feature when vowel context and speaker identity were statistically controlled. Centroid frequency yielded higher classification accuracy than Peak frequency, and features computed from psychoacoustic spectra were equally successful as those from acoustic spectra.

We used a mixed-effects logistic regression model with spectral feature and vowel context as fixed effects, and speaker identity as a grouping factor to differentiate /t/ and /k/ productions. This approach was described by Holliday et al., (2015) as a way to quantify robustness of an individual's /s/-/ʃ/ contrast. The primary objective of this type of model—one that is not independent of speaker or vowel context—is to quantify the relationship between productions of two target categories within a speaker. The model uses by-participant random effects to make individualized predictions, and ultimately the variable of interest derived from the model is the percentage of tokens correctly predicted for each speaker. This variable indexes one notion of distance between sets of productions (cf. the mean Mahalanobis distance between two sets of points, the distance between the means of two sets of points, or the discriminability between the sets of points).

Researchers and clinicians alike can use this approach to determine the extent to which two sets of productions are separable within a speaker and make comparisons over time or across groups. For example, using a similar logistic regression classifier, Nicholson et al. (2015) found that robustness of the /s/-/ʃ/ contrast increases with age and vocabulary. Todd, Edwards, and Litovsky (2011) showed that children with cochlear implants produce less robust /s/-/ʃ/ contrasts than peers with normal hearing, even when tokens are transcribed as correct. These studies

speak both to the importance of using continuous, acoustic measures to characterize productions, and to the utility of a within-participant measure of robustness of contrast. The aim of this study was to empirically compare spectral features that could differentiate /t/ and /k/ productions within and across adult speakers, and that would be accessible to a range of professionals.

There are some limitations to this classification approach. First, it requires several tokens per category per context per speaker for the mixed-effects logistic regression model to work reliably. This increases the time required to collect and code production data. However, the coding procedure was streamlined and largely automated, so each token was transcribed and tagged in approximately one minute. Second, our method of coding vowel context qualitatively as "front" or "back" based on the target word is not always feasible. The vowel could be centralized in some productions or dialects, misarticulated by children, or the target word may not be known. In these cases, an on-line coding procedure to label the vowel for each token may be necessary. Quantitative representations of formant transitions could also be incorporated, but would substantially increase the amount of time and expertise required to obtain reliable measurements. Finally, using a logistic regression model that includes an indexical grouping factor is not well suited for all classification purposes. This approach would not translate easily to applications with the goal of low-resource, fully-automatic recognition and classification of stop consonants.

Previous work (e.g., Kewley-Port, 1983; Nossair & Zahorian, 1991) suggested that dynamic, global spectral shape features are superior to static features for identifying stop consonants across speakers and contexts. McMurray and Jongman

(2011) found that no acoustic parameters were unaffected by context for fricatives. Perhaps dynamic features calculated over relatively long analysis windows, especially ones that overlap with following vowel, serve as a compensation mechanism for variability. We acknowledge the importance of dynamic cues in differentiating stop consonants, and we submit that including a categorical variable of vowel context provides a sufficient, yet simpler approach for encoding dynamic features. The success of our single, static feature (calculated purposefully from a window excluding any voicing) supports the idea that differences in energy concentration in the burst alone provide sufficient information to differentiate /t/ from /k/, when those values are considered relative to expectations for an individual speaker within a given vowel context.

Future work could: compare traditional dynamic features to the current approach and determine whether there are significant performance benefits; assess clinicians' ability and willingness to implement our classification approach compared to one that relies on dynamic features; and determine whether Centroid frequency is also sufficient for classifying children's /t/ and /k/ productions, which are notoriously more variable. Finally, it will be important to validate the current findings with a perceptual measure, as in Holliday et al., (2015). Perhaps acoustic features computed from acoustic and psychoacoustic spectra yield equivalent classification accuracy, but features from one representation better align with listeners' perceptual ratings.

Chapter 3: Effects of device limitations on acquisition of the /t/-/k/ contrast in children with cochlear implants[5]

Abstract

Objectives: The present study investigated how development of the /t/-/k/ contrast is affected by the unique perceptual constraints imposed on young children using cochlear implants (CIs). We hypothesized that children with CIs would demonstrate unique patterns of speech acquisition due to device limitations, rather than straightforward delays due to a lack of auditory input in the first year of life before implantation. This study focused on the contrast between /t/ and /k/ because it is acquired early in the sequence of development, requires less advanced motor control than later-acquired place contrasts, is differentiated by spectral cues (which are particularly degraded when processed by CIs), and is not easily differentiated by visual cues alone. Furthermore, perceptual confusability between /t/ and /k/ may be exacerbated in front-vowel contexts, where the spectral energy for /k/ is shifted to higher frequencies, creating more spectral overlap with /t/.

Design: Children with CIs (n=26; ages 31 to 66 mo) who received implants around their first birthdays were matched to peers with normal hearing (NH). Children participated in a picture-prompted auditory word-repetition task that included over 30 tokens of word-initial /t/ and /k/ consonants. Tokens were balanced across front-

[5] Reproduced from Johnson, A.A., Bentley, D.M, Munson, B., & Edwards, J. (2021). Effects of device limitations on acquisition of the /t/-/k/ contrast in children with cochlear implants. *Ear & Hearing*, Epub ahead of print, doi: 10.1097/AUD.0000000000001115, with the permission of Wolters Kluwer Health - LWW Publishing)

vowel and back-vowel contexts to assess the effects of coarticulation. Productions were transcribed and coded for accuracy as well as the types of errors produced (manner of articulation, voicing, or place of articulation errors). Centroid frequency was also calculated for /t/ and /k/ tokens that were produced correctly. Mixed-effects models were used to compare accuracy, types of errors, and centroid frequencies across groups, target consonants, and vowel contexts.

Results: Children with CIs produced /t/ and /k/ less accurately than their peers in both front- and back-vowel contexts. Children with CIs produced /t/ and /k/ with equal accuracy, and /k/ was produced less accurately in front-vowel contexts than in back-vowel contexts. When they produced errors, children with CIs were more likely to produce manner errors and less likely to produce voicing errors than children with NH. Centroid frequencies for /t/ and /k/ were similar across groups, except for /k/ in front-vowel contexts: children with NH produced /k/ in front-vowel contexts with higher centroid frequency than children with CIs, and they produced /k/ and /t/ with equal centroid frequencies in front-vowel contexts.

Conclusions: Children with CIs not only produced /t/ and /k/ less accurately than peers with NH, they also demonstrated idiosyncratic patterns of acquisition, likely resulting from receiving degraded and distorted spectral information critical for differentiating /t/ and /k/. Speech-language pathologists should consider perceptual confusability of consonants (and their allophonic variations) during their assessment and treatment of this unique population of children.

Introduction

Cochlear implants (CIs) have changed the way children born with profound hearing loss learn speech and language. Children with profound hearing loss (>90 dB Hearing Level [HL]) who use CIs often have similar speech outcomes to children with less severe degrees of hearing loss who use hearing aids (Baudonck et al. 2011; Fitzpatrick et al. 2012; Osberger, 1997), though speech acquisition is still delayed compared to peers with normal hearing (NH; Bass-Ringdahl, 2010; Baudonck et al. 2011; Blamey et al. 2001; Spencer & Guo, 2012).

There are two key factors impacting speech acquisition in children with CIs: a period of auditory deprivation early in life prior to implantation, and the limitations of current CI devices and signal processing strategies. When data were collected for this study, CIs were approved by the Federal Drug Administration (FDA) in the United States for children over 12 months of age; currently, CIs manufactured by Cochlear™ are approved for children as young as 9 months. Critical stages of speech development—including perceptual reorganization for language-specific speech sounds, and the onset of canonical babbling—occur within the first 9 months of life for children born with NH (Kuhl et al. 1992; Oller et al. 1999; Werker & Tees, 1984). Since children born with profound hearing loss in the United States must wait at least 9 months to begin effectively learning language through auditory input and perceptual-motor feedback loops, it is logical that speech development is delayed relative to peers with NH.

In addition to a period of auditory deprivation, the progression of speech development is also likely affected by device limitations. In the signal delivered by a

CI, temporal envelope information is prioritized, temporal fine structure is discarded, and spectral information is degraded (Loizou, 2006; Zeng et al. 2008). CIs distort spectral cues in several ways. First, there are fewer sites of stimulation along the tonotopically-organized auditory nerve, so the number of unique frequencies available to the listener is reduced. Second, the electrode array is shorter than the basilar membrane. With no electrodes to stimulate the apex (high frequency boundary) or the base (low frequency boundary), the overall range of frequencies available to the listener is more compact. Third, each electrode stimulates a large area of nerve endings in a uniform way throughout the cochlea. In contrast, in acoustic hearing, hair cells in a healthy cochlea are aligned with the nerve endings, which leads to finer frequency resolution than is possible in electric hearing. Fourth, in a healthy auditory system, outer hair cells provide feedback to the basilar membrane to influence its movement and increase frequency resolution, but CIs do not replicate this physiology. Finally, there are mismatches between the natural frequency response of the auditory nerve fibers being stimulated and the CI's frequency map. Consequently, speech sounds and contrasts that rely on steady-state or temporal envelope information (such as consonants differentiated by voicing cues) are transmitted relatively well, but those that rely on spectral cues (such as consonants differentiated by place cues) are substantially degraded.

Although it is clear that young children with CIs do not acquire speech sounds at the same ages as children born with NH, current research on speech acquisition does little to differentiate between delays due to early auditory deprivation and differences due to CI device limitations. Understanding the influence of device

limitations on speech acquisition is critical to informing unique approaches to speech therapy and improving speech processing strategies to facilitate perception and learning. Our study addresses this gap by examining in detail the acquisition of the contrast between /t/ and /k/, an early acquired contrast that depends primarily on spectral cues.

One study by Spencer and Guo (2012) assessed articulation skills and patterns of errors produced by 32 children who received implants by age 2;6 (2 years; 6 months) using the *Goldman-Fristoe Test of Articulation-2nd Edition* (GFTA-2; Goldman & Fristoe, 2000). They concluded that, with few exceptions, children with CIs acquire consonants in a similar fashion (in terms of both rate and order) as children with NH who have matched hearing experience. They also found that the types of errors children with CIs produced (omissions and substitutions) were similar to developmental patterns expected for children with NH. The finding that speech acquisition in children with CIs was delayed but not qualitatively different from children with NH constitutes evidence that auditory deprivation is the primary factor influencing speech acquisition for young children with CIs.

Several limitations of Spencer and Guo (2012) are worth noting. First, the speech sample included only one production of each consonant per word position. Thus, it was not possible to analyze patterns of production within a single consonant–or contrast–of interest. For example, it was not possible to determine whether there was a relationship between productions of specific consonants and the unique spectral degradations to those consonants by CI signal-processing algorithms. An in-depth speech profile (including a measure of within-speaker variation for specific

consonants) would be useful for assessing speech development in children with CIs and comparing patterns to those observed in children with NH. Second, comparisons across consonants were confounded by vowel context. For example, word-initial /t/ was elicited only in a front-vowel context, and word-initial /k/ was elicited only in a back-vowel context. Exploring the effects of vowel context is particularly important for consonants affected by coarticulation, such as /k/. Sub-phonemic variability in the acoustic signal may not be transmitted faithfully by the CI, so children may conflate allophonic and phonemic differences while learning to talk. Third, Spencer & Guo (2012) characterized speech using only transcriptions. Recent research reported quantifiable acoustic differences between consonants produced by children with CIs and children with NH, even for sounds that were transcribed as correct (Reidy et al. 2017; Todd et al. 2011). Thus, acoustic measures of children's speech to supplement transcriptions would allow for richer comparisons of speech patterns within and across children.

Research using spectral features to compare productions across groups provides some insight into how device limitations affect speech acquisition in children with CIs. Studies have focused on /s/ and /ʃ/, a place contrast with sounds differentiated by spectral features that are typically acquired after age 6 in children with NH (Smit et al. 1990) and after at least 2-4 years of device use in children with CIs (Blamey et al. 2001; Spencer & Guo, 2012). One reason /ʃ/ and especially /s/ are acquired relatively late in the sequence for children with NH is because of increased motor demands: these sounds involve precise tongue bending, positioning, and continuously controlled airflow—gestures which require substantial neuromotor

coordination. Thus, there are two potential reasons that Reidy et al. (2017) and Todd et al. (2011) found that children with CIs produced /s/ with lower spectral energy than children with NH. One reason for the difference across groups, as suggested by the authors, is that the high-frequency spectral information (above 8,000 Hz) used to differentiate /s/ is eliminated by the CI processor, so children learn to produce /s/ with lower centroid frequency than their peers. Alternatively, differences could be due to maturing motor systems: children typically master /s/ later than /ʃ/ because more advanced motor skills are required to produce /s/ (Smit et al. 1990). Studying acquisition of the contrast between /t/ and /k/ may help clarify the effects of perceptual constraints without the influence of production constraints. This contrast is also differentiated by spectral features, but it is acquired earlier when motor control is less advanced. Furthermore, /t/ is typically acquired before /k/ in children with NH (Smit et al. 1990), but the release burst for /t/ has high-frequency components that extend beyond the 8,000 Hz upper limit of the CI range, more similar to /s/.

There are several other ways device limitations affect perception of the English /t/-/k/ place contrast. These stop consonants have short duration, transient acoustic cues, low intensity, high-frequency spectral components, several allophonic variations, and they are differentiated by spectral cues by children with NH. Furthermore, the acoustic signal of /k/ is transformed by vowel context: the spectral energy in the release burst for /k/ shifts to higher frequencies when followed by a front vowel. Thus, the spectral energy in /k/ before a front vowel overlaps considerably more with the release burst for /t/. This sub-phonemic difference does not appear to affect learning in children with NH; however, the increased spectral

overlap likely exacerbates perceptual confusability for children using CIs who have poor frequency resolution.

Children with CIs experience an interval of auditory depravation prior to implantation. Hence, it is relevant to consider the error patterns of /t/ and /k/ by children with hearing loss who communicate orally and those who use hearing aids. One representative study of these patterns is presented by Smith (1975), who examined the speech of 40 children (8 to 15 years) who were Deaf or Hard of Hearing, used hearing aids, and attended oral schools. The overall accuracy of these children's productions of /t/ and /k/ were low: across word positions, only 40% of children's /k/ productions and 37% of children's /t/ productions were accurate. The most common error pattern was omissions (51% for /t/ and 46% for /k/ in initial position), and the most common substitution error was a glottal stop for both /t/ and /k/. As will be evident below, this pattern is very different from what we observed in this study, even with children who are considerably younger than in the Smith (1975) study.

The present study aimed to extend our understanding of speech acquisition in young children with CIs by examining the influence of device limitations. We compared overall accuracy, patterns of speech errors, and spectral features of /t/ and /k/ sounds produced by 3- to 5-year-old children with CIs and peers with NH. Unlike previous research, the current study elicited numerous /t/ and /k/ tokens in initial position from each participant using a standardized word repetition task. The current study also balanced tokens across front- and back-vowel contexts to directly assess

the effects of coarticulation, and analyzed productions using both transcription and acoustic measures.

If auditory deprivation alone accounted for differences in acquisition between groups, children with CIs would show similar developmental patterns as children with NH, just on a delayed timeline. In this case, children with CIs in the current study would produce /t/ and /k/ less accurately overall than peers with NH, and /t/ would be produced more accurately than /k/ during development (before children have mastered both sounds). Children in the current study would also produce the same types of developmental errors that children with NH produce, such as prevocalic voicing (e.g., "d" for /t/) and velar fronting (e.g., "t" for /k/) (e.g., McLeod, & Bleile, 2003).

Alternatively, if device limitations accounted for differences across groups, children with CIs would demonstrate unique patterns of acquisition and produce errors that are not typical for children with NH. We hypothesized that, due to degraded spectral cues and perceptual confusability, children with CIs would produce /t/ and /k/ less accurately than peers with NH, /t/ and /k/ would be produced with equal accuracy. We further hypothesized that degraded spectral cues would lead to particularly low production accuracy in front-vowel contexts. Coarticulation increases the centroid frequency of /k/ bursts in front-vowel contexts, leading to more spectral overlap with /t/ bursts. Thus, the contrast between /t/ and /k/ is spectrally less robust in front-vowel contexts compared to back-vowel contexts (e.g., Johnson et al., 2018). The decreased frequency resolution imposed by CI signal-processing algorithms is predicted to compromise the subtler spectral cues in front-vowel contexts while

somewhat preserving the more robust spectral cues between /t/ and /k/ in back-vowel contexts. In terms of specific error patterns, we hypothesized that children with NH would produce developmental errors such as prevocalic voicing and velar fronting, but children with CIs would produce relatively few voicing errors (because of access to salient temporal cues) and more place errors (because of degraded spectral cues). Finally, we hypothesized that children with CIs would produce /t/ with lower spectral energy than children with NH, similar to the trend for /s/ productions reported by Reidy et al. (2017) and Todd et al. (2011).

Materials and Methods

This study was approved by the IRBs at the University of Wisconsin, Madison and the University of Minnesota, Twin Cities where data were collected and coded, and by the IRB at the University of Maryland, College Park, where data were analyzed.

Participants

Twenty-eight children with CIs and 28 children with NH were recruited to participate in this study. Some children (n=11) were part of a larger longitudinal project and returned to the laboratory once per year over a 2-to-3-year period. Every child with a CI was matched to a child with NH in terms of age, sex, and maternal education. Two of the children did not have suitable matches, so 26 children (15 females, 11 males) were included in each group. None of the children had disabilities

per parent report. Eleven matched pairs participated more than once over the course of three years. Each visit was treated as a unique observation, but non-independence due to multiple visits for some children was accounted for in the random effects structure of the mixed effects models used for statistical analyses. There were 15 children from each group who participated in 1 session, 8 children who returned 1 year later, and 3 children who participated 3 times. Thus, there were 40 sessions of data collected for each group. Demographic information for each group is presented in Table 1.

Table 1. Participant demographic information.

Group	Age	Boys/ Girls	*EVT-2*[a] Standard score (*SD*): 100 (*15*)	*PPVT-4*[b] Standard score (*SD*): 100 (*15*)	*GFTA-2*[c] Standard Score (*SD*): 100 (*15*)
CI	50 months range: 31-66 n = 40	16/ 24	98 (*19*) range: 46 – 131 n = 39	94 (*21*) range: 40 – 139 n = 39	74 (*20*) range: 39 – 107 n = 36
NH	50 months range: 32-66 n = 40	16/ 24	118 (*11*) range: 88 – 134 n = 40	121 (*11*) range: 94 – 140 n = 22	92 (*12*) range: 67 – 113 n = 26

[a] *Expressive Vocabulary Test-2nd Edition* (Williams, 2007)
[b] *Peabody Picture Vocabulary Test-4th Edition* (Dunn & Dunn, 2007)
[c] *Goldman-Fristoe Test of Articulation-2nd Edition* (Goldman & Fristoe, 2000)

Participants' mothers in each group had identical education levels, including graduate degrees (n = 7), 4-year college degrees (n = 12), some college, trade school, or technical/associate degrees (n = 5), and high school diploma or some high school (n = 2). For children with CIs, hearing loss was identified at a mean age of 4.1

months (*SD* = 7.1; median = 0; range: 0-30), and the mean age of activation was 17.9 months (*SD* = 10.3; median = 14; range: 6-45). Nineteen children used bilateral CIs, 5 children used a bimodal system (CI in one ear and a hearing aid in the other), and 2 children used only a unilateral CI. Twenty-three children used hearing aids prior to implantation, and information about pre-implant amplification was not available for the other three children. All children with CIs participated in the *Ling-6 Sound Test* (Ling, 1976; 1989) to ensure the implants were functioning properly; one child who did not pass the test had their implant re-mapped (and re-tested) by a certified audiologist before participating further.

Stimuli

Children participated in a picture-prompted auditory word-repetition task. Stimuli included 32 different familiar words (17 word-initial /k/ targets, 15 word-initial /t/ targets) spoken in isolation by an adult female. To ensure familiarity, words that 90% of children understand and produce were chosen (according to the *Macarthur-Bates Communication Development Inventory* norms [Fenson et al., 2007]). The complete list of /t/ and /k/ words can be found in Appendix A (Supplemental Digital Content 1). Wordlists also included a variety of filler words that did not begin with /t/ or /k/. Recordings were made in a sound-treated recording booth and amplitude normalized to 70 dB. A high-quality archetypal photograph, edited for consistency in size and background, was paired with each word. The 32 different words were distributed across three wordlists, each intended for children within a specified age range. For the youngest age group, there were 9 different /k/-

initial words and 8 different /t/-initial words, half in front-vowel context (followed by /i ɪ e ɛ æ/) and half in back-vowel context (followed by /u ʊ o ʌ ɑ/ or the diphthongs /aɪ aʊ/). Each word was repeated twice to elicit 34 productions per child. As children got older, the number of different words on the wordlist increased (because new words became familiar and available for use), and the number of repeated words decreased. Tokens were always balanced across front- and back-vowel contexts within a wordlist.

Procedure

All children with NH passed a hearing screening and all children with CIs passed the *Ling-6 Sound Test* (Ling, 1976; 1989) before participating in any other tasks. Testing took place in a sound-treated recording booth. During the experimental task, children sat in front of a computer screen positioned approximately six inches away from a Shure SM81 cardioid condenser microphone with a custom pop filter. Words were presented over loudspeakers at a comfortably loud listening level in a pseudo-randomized order, with steps taken to ensure target words were not repeated on consecutive trials. Picture stimuli appeared simultaneously with auditory stimuli. After word offset, children repeated the word into the microphone, and an experimenter recorded the session using a Marantz PMD671 solid-state recorder at a sampling frequency of 44,100 Hz. Words could be repeated once if the child didn't hear the first presentation due to attention or noise. Children were asked to repeat themselves if a response was obscured by background noise (e.g., the child responded while touching the microphone, tapping the table, kicking the chair), or if the

recording clipped because the child spoke significantly louder than during calibration. During their visits to the laboratory, children completed other experimental tasks and standardized tests as part of a larger project, and parents filled out questionnaires about their child's developmental history and family demographics.

Coding

Coding was done in *Praat* (Boersma & Weenink, 2018). Two trained phoneticians (the first and second authors) coded manner, place, and voicing for all target /t/ and /k/ tokens. Options for manner transcriptions included [Stop], [Affricate], and [Other]. The [Other] category included fricative substitutions, approximant substitutions, deletions, distortions, and other articulations that were not easily transcribed, but were decidedly neither [Stop] nor [Affricate] manners of articulation. Options for place transcriptions included [t], [k], intermediate categories of [t:k] (intermediate, closer to [t]) and [k:t] (intermediate, closer to [k]), and [other] (bilabial or glottal stops). The use of intermediate categories was motivated by Stoel-Gammon's (2001) recommendation to code 'fuzzy' categories that are not clear exemplars of sound categories. For productions transcribed as [Stops], the coders marked locations on the waveform corresponding to the release burst (first transient-noise spike following a period of silence that coincided perceptually with the release of an oral constriction) and the onset of voicing (first upward swing from the zero-crossing followed by a stable, quasi-periodic pattern of voicing). Consonants were coded as 'voiceless' if the voice onset time (VOT; time between release burst and

onset of voicing) was calculated to be greater than 20ms. Neither place nor voicing were coded for productions with manner transcriptions other than [Stop].

Productions were scored as Correct (i.e., Accuracy = 1) when they were transcribed as a [Stop], had VOT >20 ms, and had a place transcription of [t] or [t:k] for target /t/ or [k] or [k:t] for target /k/. Intermediate tokens were coded as correct if they were closer to the target consonant, because these productions would have been categorized as the target using standard transcription. Incorrect productions were coded as: Manner errors (i.e., Manner error = 1) when the manner of articulation was coded as anything other than [Stop]; Place errors (i.e., Place error = 1) when the manner of articulation was coded as [Stop], but the place of articulation did not match the target consonant (e.g., "t" for /k/); or Voicing errors (i.e., Voicing error = 1) when the manner of articulation was [Stop], but VOT was <20 ms (e.g., "d" for /t/). Productions transcribed as [Stop] could have both Place and Voicing errors (e.g., "g" for /t/).

A random 20% of the files were coded by both researchers for reliability purposes. Agreement between the two transcribers on whether the sound was produced accurately was 95% for children with CIs and 97% for children with NH.

Spectral Measures

For tokens that were transcribed as correct and had release bursts unobstructed by noise, acoustic spectra were estimated using custom scripts in the *R* programming environment (R Core Team, 2013). Centroid frequencies were calculated using the procedures reported in detail by Reidy (2016) and validated for /t/ and /k/ in Johnson

et al. (2018). To summarize: for each token, a 25-ms analysis window was defined from 5 ms prior to the release burst through 20 ms after the burst tag. Within this 25-ms window, the acoustic spectrum of the waveform was estimated using an eighth-order multitaper spectrum. The values of each spectrum were normalized so they summed to 1, and the spectrum was treated as a probability mass function over frequency. The Centroid frequency for each token was the mean value of this distribution.

Accuracy Analysis

Four logistic mixed-effects regression models with homologous structures were fit using the *R* lme4 package (Bates et al. 2014, v. 1.1-11). Each model predicted the log-likelihood that a token was produced accurately based on: fixed effects of Group (β_1; CI or NH), Target Consonant (β_2; /t/ or /k/), Vowel Context (β_3; front or back), the 2-way interactions between Group and Target Consonant (β_4), Group and Vowel Context (β_5), and Target Consonant and Vowel Context (β_6), the 3-way interaction between Group, Target Consonant, and Vowel Context (β_7), as well as the participant-level random intercept (u_0) and random slope for Target Consonant (u_1). Each model was fit with a different reference category in order to compare accuracy within and across groups, consonants, and contexts. The structure of Models 1a-1d is formalized in Equation 1, with subscripts *i* and *j* representing group- and participant-level variables, respectively. (For more information on mixed-effects modeling, see Brauer & Curtain, 2017; Jaeger, 2008.) To account for fitting 4 models to the same data, an adjusted alpha-level of $p = 0.0125$ was used to evaluate significance.

Equation 1.

$$log\left(\frac{Accuracy}{1-Accuracy}\right)_{ij} = \beta_0 + \beta_1 \times Group_j + \beta_2 \times Target\ Consonant_{ij} + \beta_3 \times Vowel\ Context_{ij} + \beta_4 \times Group_j \times Target\ Consonant_{ij} + \beta_5 \times Group_j \times Vowel\ Context_{ij} + \beta_6 \times Target\ Consonant_{ij} \times Vowel\ Context_{ij} + \beta_7 \times Group_j \times Target\ Consonant_{ij} \times Vowel\ Context_{ij} + u_{0j} + u_{1j} \times Target\ Consonant_{ij} + \varepsilon_{ij}$$

Error Pattern Analysis

Six logistic mixed-effects regression models with homologous structures (2 for each type of error with different reference categories) were used to predict the log-likelihood that an incorrect token contained a given type of error based on: fixed effects of Group, Target Consonant, and the 2-way interaction between Group and Target Consonant, as well as the participant-level random intercept. The structure of Models 2a and 2b is formalized in Equation 2, with subscripts *i* and *j* representing group- and participant-level variables, respectively.

Equation 2.

$$log\left(\frac{Manner\ Error}{1-Manner\ Error}\right)_{ij} = \beta_0 + \beta_1 \times Group_j + \beta_2 \times Target\ Consonant_{ij} + \beta_3 \times Group_j \times Target\ Consonant_{ij} + u_{0j} + \varepsilon_{ij}$$

Models 2c-2f had homologous structures but predicted the log-likelihood that an incorrect token was produced with a Voicing error (Models 2c-2d) or a Place error (Models 2e-2f). By-participant random slopes for Target Consonant were not

appropriate for these models, because many children produced errors on either /t/ or /k/, but not both, so Target Consonant did not consistently vary within-participant. Vowel Context was excluded from the error pattern analyses, because there were not enough data from each category. To account for fitting 2 models to the same data, an adjusted alpha-level of $p = 0.025$ was used to evaluate significance.

Spectral Feature Analysis

Four linear mixed-effects regression models with homologous structures (but different reference categories) predicted Centroid frequency (kHz) based on: fixed effects of Group, Target Consonant, Vowel Context, and all interactions, as well as the participant-level random intercept and random slope for Target Consonant. The model equations were identical to that formalized in Equation 1, except the dependent variable was $Centroid_{ij}$. Models 3a-3d had the same reference categories as Models 1a-1d, respectively, and an adjusted alpha-level of $p = 0.0125$ was used to evaluate significance.

Results[6]

Accuracy Results

There were 2543 productions transcribed and included in the accuracy analysis (some tokens could not be transcribed because the child did not produce the target or produced an inaudible response). On average, children produced 32 ($SD = 3$)

[6] Detailed model results for all analyses included in this chapter are included in Appendix C; original text retained.

transcribable tokens per session. Children with CIs produced target /t/ with 70% accuracy (*SD* = 46%) and target /k/ with 65% accuracy (*SD* = 48%). Children with NH produced target /t/ with 93% accuracy (*SD* = 26%), and target /k/ with 92% accuracy (*SD* = 27%). Table 2 shows the percentage of tokens produced correctly for each group by target consonant and vowel context. The results of Models 1a-1d are presented in Appendices B, C, D, and E (Supplemental Digital Content 2).

Table 2. The mean percentage (and *SD*) of /t/ and /k/ tokens produced correctly by children with CIs and children with NH in front- and back-vowel contexts.

	/t/		**/k/**	
	Front-vowel context	**Back-vowel context**	**Front-vowel context**	**Back-vowel context**
CI	69% (46%)	70% (46%)	60% (49%)	70% (46%)
NH	91% (29%)	94% (24%)	93% (26%)	92% (28%)

Model 1a (reference: CI, target /k/, back-vowel context) showed significant main effects of intercept ($\hat{\beta}_0$ = 1.22, *SE* = 0.35, *z* = 3.45, *p* < 0.001), Group ($\hat{\beta}_1$ = 2.11, *SE* = 0.54, *z* = 3.87, *p* < 0.001), Vowel Context ($\hat{\beta}_3$ = -0.68, *SE* = 0.23, *z* = -3.00, *p* = 0.003), and a significant 3-way interaction ($\hat{\beta}_7$ = -1.55, *SE* = 0.59, *z* = -2.62, *p* = 0.009). The interaction between Group and Vowel Context ($\hat{\beta}_5$ = 0.95, *SE* = 0.42, *z* = 2.28, *p* = 0.02) and between Target Consonant and Vowel Context ($\hat{\beta}_6$ = 0.71, *SE* = 0.32, *z* = 2.23, *p* = 0.03) did not reach significance after the adjusted alpha-level was applied.

Model 1b (reference: CI, target /t/, front-vowel context) showed novel significant main effects of intercept ($\hat{\beta}_0$ = 1.15, *SE* = 0.32, *z* = 3.57, *p* < 0.001) and Group ($\hat{\beta}_1$ = 1.94, *SE* = 0.50, *z* = 3.90, *p* < 0.001). The interaction between Target Consonant and Vowel Context is identical to Model 1a, and the significant 3-way interaction is identical across all four models. The main effect of Target Consonant ($\hat{\beta}_2$ = -0.61, *SE* = 0.30, *z* = -2.06, *p* = 0.04) and the interaction between Group and Target Consonant ($\hat{\beta}_4$ = 1.11, *SE* = 0.52, *z* = 2.12, *p* = 0.03) did not reach significance once the adjusted alpha-level was applied.

Model 1c (reference: NH, target /k/, front-vowel context) showed novel significant main effects of intercept ($\hat{\beta}_0$ = 3.59, *SE* = 0.43, *z* = 8.32, *p* < 0.001) and Group ($\hat{\beta}_1$ = -3.05, *SE* = 0.55, *z* = -5.54, *p* < 0.001). The interaction between Group and Target Consonant is identical to Model 1b, and between Group and Vowel Context is identical to Model 1a.

Model 1d (reference: NH, target /t/, back-vowel context) showed novel significant main effects of intercept ($\hat{\beta}_0$ = 3.66, *SE* = 0.41, *z* = 8.86, *p* < 0.001) and Group ($\hat{\beta}_1$ = -2.55, *SE* = 0.52, *z* = -4.92, *p* < 0.001). The interaction between Group and Target Consonant is identical to Model 1a, between Group and Vowel Context is identical to Model 1b, and between Target Consonant and Vowel Context is identical to Model 1c.

Taken together, the significant main effects of Group in all four models indicates that children with CIs produced both /t/ and /k/ less accurately than peers

with NH in both front- and back-vowel contexts. Accuracy across groups is illustrated in Figure 1.

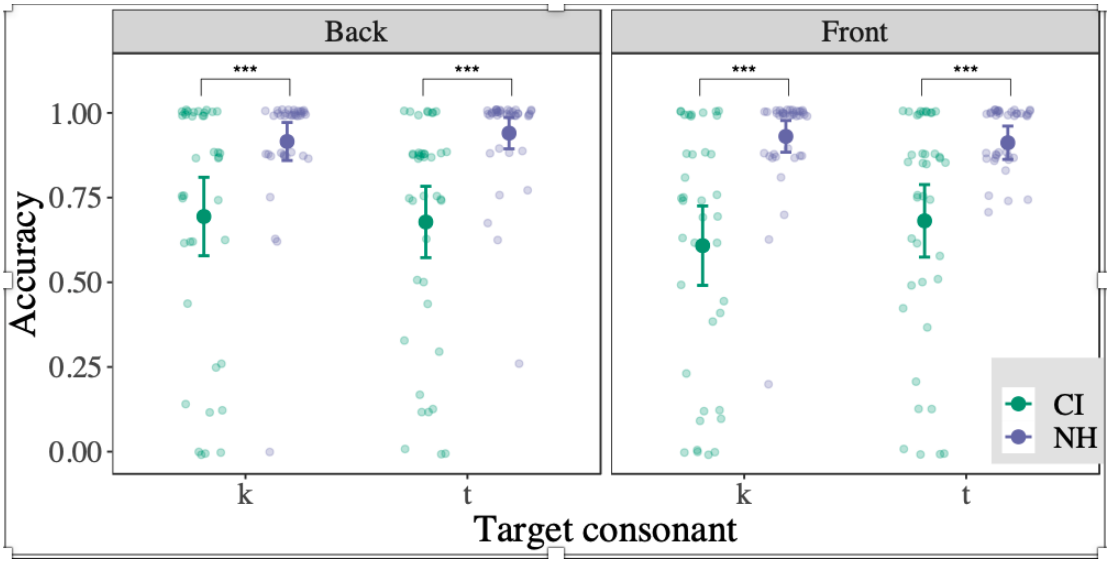


Figure 1. Mean accuracy scores for each group (large circles) with ±2 standard error bars and individual data (smaller transparent circles) for word-initial /k/ and /t/ tokens in both back- and front-vowel contexts. *** indicates $p < 0.001$

The non-significant main effects of Target Consonant in all four models indicate that children in each group produced /k/ as accurately as /t/ within each vowel context; however, the trend in Model 1b suggests children with CIs produce /k/ somewhat less accurately than /t/ in front-vowel contexts. The significant main effect of Vowel Context in Model 1a suggests that children with CIs produced /k/ less accurately in front-vowel contexts compared to /k/ in back-vowel contexts, even though vowel context did not affect accuracy for /t/ (Model 1b), nor did it affect accuracy of either consonant for children with NH (Models 1c and 1d). The complicated relationship between accuracy, target consonant, and vowel context is further elucidated by the significant 3-way-interaction between Group, Vowel Context, and Target Consonant. The 3-way interaction indicates that the difference in

accuracy between /t/ and /k/ across vowel contexts was larger for children with CIs compared to children with NH (who were close to ceiling-level accuracy across all contexts). Children with NH produced /t/ and /k/ with equal accuracy regardless of vowel context, whereas children with CIs produced /t/ and /k/ with more similar accuracy in back-vowel contexts compared to front-vowel contexts. The greater difference in accuracy across vowel contexts was due to lower accuracy of /k/ in front-vowel contexts: children with CIs produced /t/ with equal accuracy across vowel contexts, but they produced /k/ less accurately in front-vowel contexts compared to back-vowel contexts. Thus, front-vowel context decreased accuracy of target /k/ to a greater extent than for target /t/ for children with CIs but did not affect accuracy of either target for children with NH.

Error Pattern Results

There were 510 tokens produced incorrectly (411 tokens from children with CIs, 99 from children with NH) included in the error pattern analysis. In terms of raw numbers, children with CIs produced more errors than children with NH in all categories. For each Group and Target Consonant, the number of occurrences of each type of error is presented in Table 3, along with the percentage of incorrect productions that contained that type of error. Percentages exceed 100% because some incorrect productions contained both voicing and place errors.

Table 3. The number (and percentage) of incorrectly produced /t/ and /k/ targets containing Manner, Voicing, and Place errors for children with CIs and children with NH.

Group	Target /t/	Target /k/	No. of Tokens (% of Total Errors*)
Manner Errors			
CI	73	59	132 (32%)
NH	1	4	5 (5%)
Voicing Errors			
CI	62	55	117 (29%)
NH	36	28	64 (65%)
Place Errors			
CI	69	145	214 (52%)
NH	11	22	33 (33%)

* Percentages within each group do not add to 100%, because 52 tokens (13%) from children with CIs and 3 tokens (3%) from children with NH contained both Voicing and Place errors.

Within incorrect productions, the likelihood that a certain type of error occurred was different across groups. For children with CIs, Place errors were the most common type of error, followed by Manner errors, then Voicing errors. For children with NH, Voicing errors were the most common, followed by Place errors, then Manner errors. Furthermore, children with NH rarely produced multiple types of errors (e.g., a child who produced place errors is unlikely to also produce voicing or manner errors), whereas the children with CIs more often produced multiple types of errors. This trend can be visualized in Figures 2-4: children who produced only one type of error (or no errors) are clustered at 1.00 (or 0.00), where the children who produced multiple types of errors are distributed throughout the range of likelihoods. The results of Models 2a-2f, predicting the log-likelihood that an incorrect production contained a Manner error (2a-2b), Voicing Error (2c-2d), or Place error (2e-2f) based on Group, Target Consonant, and the interaction between Group and Target Consonant are presented in Appendices F, G, H, I, J, and K (Supplemental Digital Content 3).

Manner errors included affricate substitutions, deletions, distortions, and other types of manner substitutions (e.g., fricative substitutions, approximant substitutions). Only affricate substitutions were coded separately and analyzed. Children with CIs produced 34 affricate substitutions (26% of the 132 manner errors), which occurred on both target /t/ and target /k/. Manner errors were extremely rare for children with NH (5 total out of 99 incorrect productions) and only included 1 affricate substitution (on target /t/). Predicting the log-likelihood that an incorrect production contained a

Manner error, Model 2a (reference: CI, target /t/) showed significant main effects of intercept ($\hat{\beta}_0$ = -0.64, *SE* = 0.24, *z* = -2.64, *p* = 0.008), Group ($\hat{\beta}_1$ = -3.43, *SE* = 1.09, *z* = -3.16, *p* =0.002), and Target Consonant ($\hat{\beta}_2$ = -0.73, *SE* = 0.24, *z* = -3.05, *p* = 0.002). Model 2b (reference: NH, target /k/) showed a significant main effect of intercept ($\hat{\beta}_0$ = -2.57, *SE* = 0.59, *z* = -4.32, *p* <0.001). Group comparisons for Manner errors are illustrated in Figure 2. These results suggest that *when* children produced errors, children with CIs were more likely to produce manner errors than children with NH when the target was /t/, though there was no difference between groups for target /k/. Children with CIs were also less likely to produce a manner error for target /k/ compared to target /t/, though manner errors were equally rare across both consonants for children with NH.

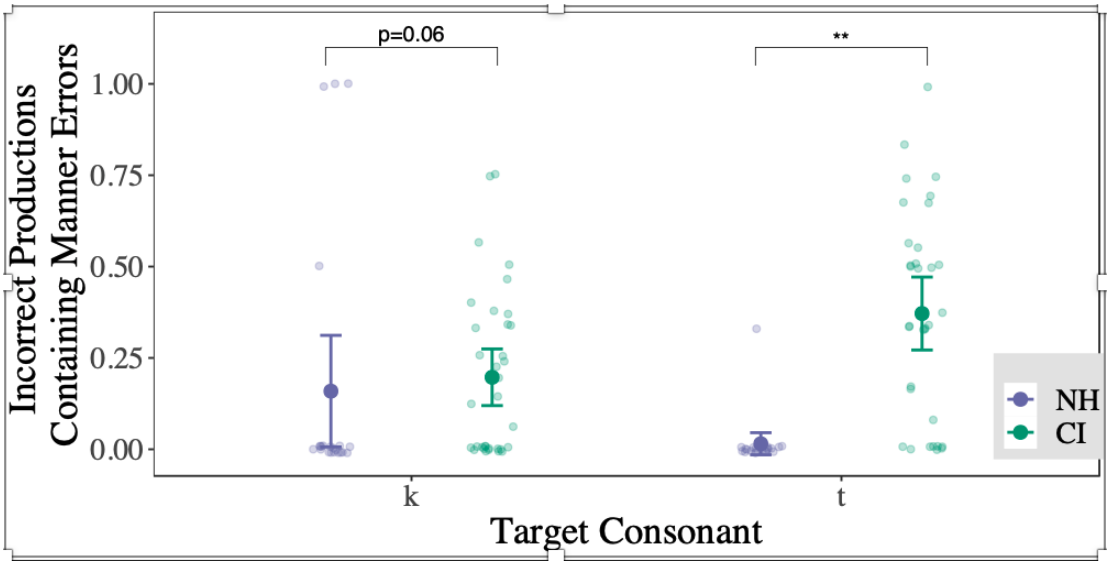


Figure 2. Mean proportion of incorrect productions that contained Manner errors for each group (large circles) with ±2 standard error bars and individual data (small circles) for word-initial /k/ and /t/ tokens in both back- and front-vowel contexts. ** indicates $p < 0.01$

Predicting the log-likelihood that an incorrect production contained a Voicing error, Model 2c (reference: CI, target /t/) showed significant main effects of intercept ($\hat{\beta}_0$ = -1.87, *SE* = 0.46, *z* = -4.05, *p* < 0.001) and Group ($\hat{\beta}_1$ = 2.86, *SE* = 0.79, *z* = 3.63, *p* < 0.001). Model 2d (reference: NH, target /k/) showed a significant main effect of Target Consonant ($\hat{\beta}_2$ = 1.92, *SE* = 0.73, *z* = 2.62, *p* = 0.009). The 2-way interaction between Group and Target Consonant did not reach significance after the adjusted alpha-level was applied ($\hat{\beta}_3$ = -1.79, *SE* = 0.79, *z* = -2.26, *p* = 0.02). Group comparisons for Voicing errors are illustrated in Figure 3. These results suggest that *when* children produced errors, children with CIs were less likely to produce voicing errors than children with NH when the target was /t/, though there was no difference between groups for target /k/. Children with NH were more likely to produce a voicing error for target /t/ compared to target /k/, though voicing errors were equally likely across consonants for children with CIs. It is important to note that the majority of voicing errors (40/64) were produced by a single child, and the majority of those errors (28/40) were produced at age 3;0. However, re-running the analysis without this child did not change our results.

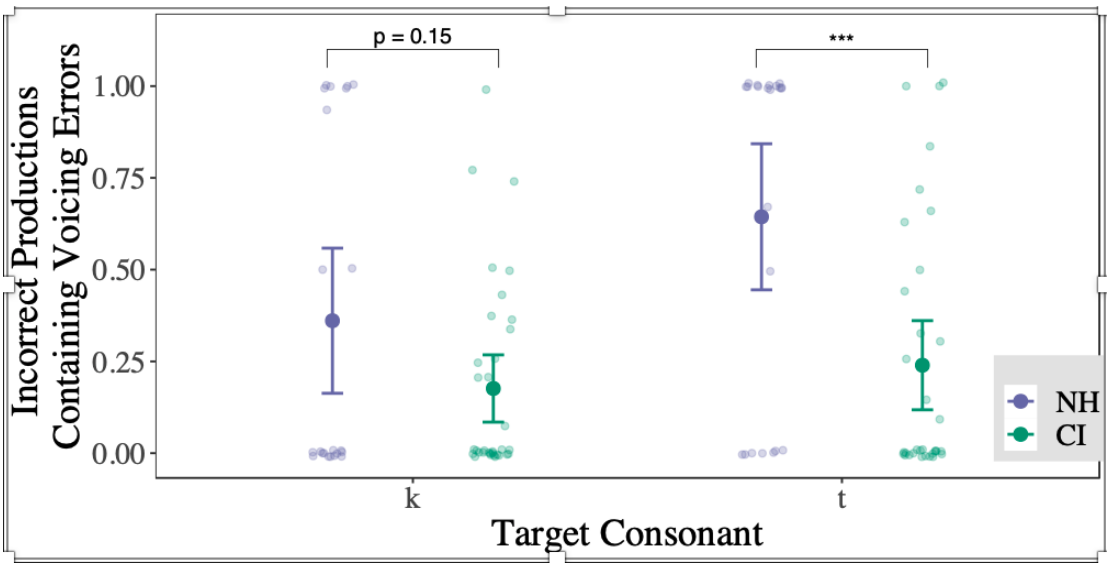


Figure 3. Mean proportion of incorrect productions that contained Voicing errors for each group (large circles) with ±2 standard error bars and individual data (small circles) for word-initial /k/ and /t/ tokens in both back- and front-vowel contexts. *** indicates $p < 0.001$

Predicting the log-likelihood that an incorrect production contained a Place error, Model 2e (reference: CI, target /t/) showed a significant main effect of Target Consonant ($\hat{\beta}_2 = 1.09$, $SE = 0.23$, $z = 4.80$, $p < 0.001$). No effects were significant in Model 2f (reference: NH, target /k/). The main effect of Target Consonant did not reach significance after the adjusted alpha-level was applied ($\hat{\beta}_2 = -1.01$, $SE = 0.51$, $z = -2.00$ $p = 0.046$). Group comparisons for Place errors are illustrated in Figure 4. These results suggest that when children produced errors, children with CIs were equally likely to produce place errors as children with NH. For children with CIs, place errors were more likely to occur for target /k/ than target /t/, and this trend was similar (though not significant) for children with NH.

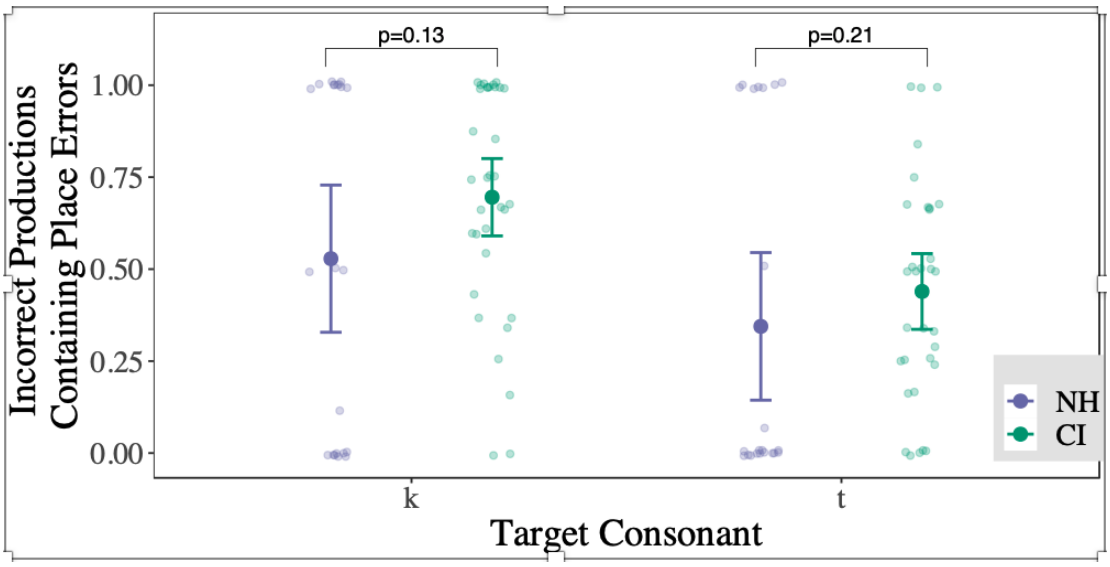


Figure 4. Mean proportion of incorrect productions that contained Place errors for each group (large circles) with ±2 standard error bars and individual data (small circles) for word-initial /k/ and /t/ tokens in both back- and front-vowel contexts.

Spectral Feature Results

Spectral measures were only calculated for /t/ and /k/ tokens that were produced correctly. There were 1945 tokens analyzed spectrally (806 correct tokens from children with CIs, 1139 from children with NH). Table 4 shows the centroid values (and standard deviations) for each group by target consonant and vowel context. The results of Models 3a-3d are presented in Appendices L, M, N, and O (Supplemental Digital Content 4).

Table 4. Centroid values (and *SD*) of /t/ and /k/ tokens produced correctly by children with CIs and children with NH in front- and back-vowel contexts.

	/t/		/k/	
	Front-vowel context	**Back-vowel context**	**Front-vowel context**	**Back-vowel context**
CI	4.18 kHz (1.36 kHz)	3.83 kHz (1.44 kHz)	3.49 kHz (0.80 kHz)	1.79 kHz (0.46 kHz)
NH	4.27 kHz (1.35 kHz)	3.84 kHz (1.23 kHz)	3.97 kHz (0.83 kHz)	1.99 kHz (0.56 kHz)

Model 3a (reference: CI, target /k/, back-vowel context) showed significant main effects of intercept ($\hat{\beta}_0$ = 1.79, *SE* = 0.07, *t* = 25.81, *p* < 0.001), Target Consonant ($\hat{\beta}_2$ = 1.97, *SE* = 0.15, *t* = 12.77, *p* < 0.001), and Vowel Context ($\hat{\beta}_3$ = 1.69, *SE* = 0.09, *t* = 19.61, *p* < 0.001), a significant interaction between Group and Vowel Context ($\hat{\beta}_5$ = 0.29, *SE* = 0.11, *t* = 2.62, *p* = 0.009), and a significant interaction between Target Consonant and Vowel Context ($\hat{\beta}_6$ = -1.36, *SE* = 0.12, *t* = -11.16, *p* < 0.001). The main effect of Group ($\hat{\beta}_1$ = 0.21, *SE* = 0.09, *t* = 2.26, *p* = 0.025) did not reach significance once the adjusted alpha-level was applied.

Model 3b (reference: CI, target /t/, front-vowel context) showed novel significant main effects of intercept ($\hat{\beta}_0$ = 4.09, *SE* = 0.16, *t* = 26.35, *p* < 0.001), Target Consonant ($\hat{\beta}_2$ = -0.61, *SE* = 0.16, *t* = -3.93, *p* < 0.001), and Vowel Context ($\hat{\beta}_3$ = -0.33, *SE* = 0.09, *t* = -3.83, *p* < 0.001). The 2-way interaction between Target

Consonant and Vowel Context is identical to Model 3a, and the 3-way interaction is identical across all four models.

Model 3c (reference: NH, target /k/, front-vowel context) showed novel significant main effects of intercept ($\hat{\beta}_0$ = 3.98, *SE* = 0.06, *t* = 66.75, *p* < 0.001), Group ($\hat{\beta}_1$ = -0.50, *SE* = 0.09, *t* = -5.35, *p* < 0.001), Vowel Context ($\hat{\beta}_3$ = -1.98, *SE* = 0.07, *t* = -27.91, *p* < 0.001), and a significant 2-way interaction between Target Consonant and Vowel Context ($\hat{\beta}_6$ = 1.51 *SE* = 0.10, *t* = 14.88, *p* < 0.001). The 2-way interaction between Group and Target Consonant is identical to Model 3b, and the interaction between Group and Vowel Context is identical to Model 3a.

Model 3d (reference: NH, target /t/, back-vowel context) showed novel significant main effects of intercept ($\hat{\beta}_0$ = 3.82, *SE* = 0.14, *t* = 26.55, *p* < 0.001), Target Consonant ($\hat{\beta}_2$ = -1.82, *SE* = 0.14, *t* = -12.89, *p* < 0.001), and Vowel Context ($\hat{\beta}_3$ = 0.46, *SE* = 0.07, *t* = 6.36, *p* < 0.001). The interaction between Group and Target Consonant is identical to Model 3a, between Group and Vowel Context is identical to Model 3b, and between Target Consonant and Vowel Context is identical to Model 3c.

Taken together, the significant main effect of Vowel Context across all four models suggests that all children produced both /t/ and /k/ consonants with higher centroid frequencies in front-vowel contexts compared to back-vowel contexts. The significant main effect of Target Consonant in Models 3a-3b suggest that children with CIs produced /k/ with lower centroid frequency than /t/ in both front- and back-vowel contexts. The significant main effect of Target Consonant in Model 3d

suggests that children with NH also produced /k/ with lower centroid than /t/ in back-vowel contexts; however, in front-vowel contexts, children with NH produce /t/ and /k/ with equal centroid frequency (Model 3c). The main effect of Group in Model 3c suggests that in front-vowel contexts, children with CIs produced /k/ with significantly lower centroid compared to children with NH, even though there was no difference between the groups for /k/ in back-vowel contexts (Model 3a) or for /t/ in either vowel context (Models 3b and 3d). Group comparisons for centroid are illustrated in Figure 5.

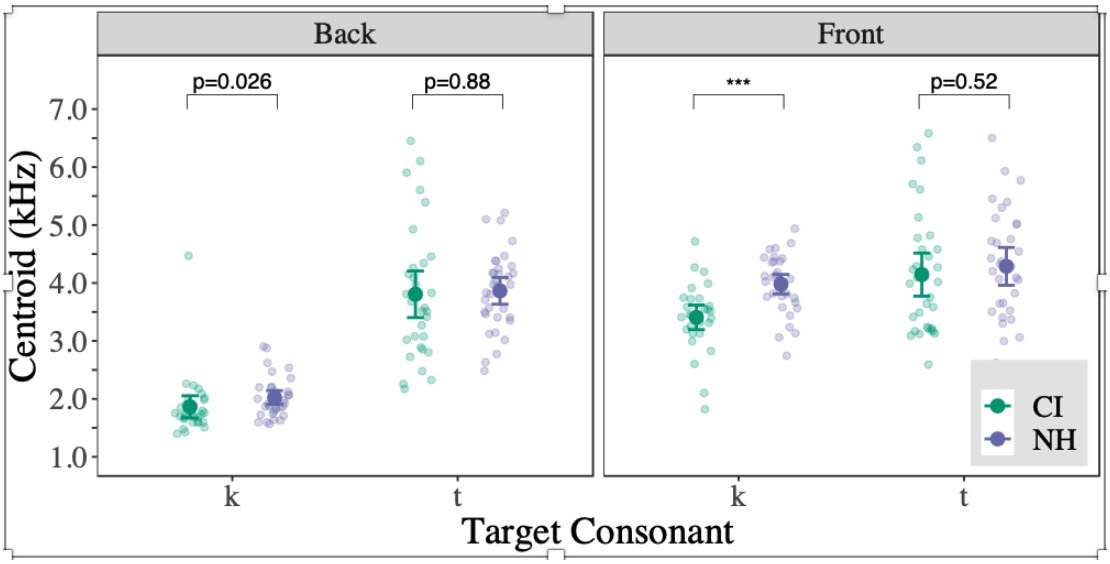


Figure 5. Mean centroid frequencies for each group (large circles) with ±2 standard error bars and individual data (smaller transparent circles) for word-initial /k/ and /t/ tokens in both back- and front-vowel contexts. *** indicates $p < 0.001$

The significant interactions between Target Consonant and Vowel Context in Models 3a-3d suggest that for both groups of children, the difference in centroid between /t/ and /k/ was larger in back-vowel contexts compared to front-vowel contexts, and the difference in centroid for /t/ across vowel contexts was smaller than

the difference for /k/ across vowel contexts. The effect is even more pronounced for children with NH: while children with CIs produced a measurable difference between /t/ and /k/ in front-vowel contexts, children with NH did not. The interaction between Group and Vowel Context in Models 3a and 3c suggest that the difference between /k/ in back-vowel contexts and /k/ in front-vowel contexts is smaller for children with CIs compared to children with NH.

Discussion

The aim of this paper was to compare accuracy, error patterns, and spectral features of word-initial /t/ and /k/ produced by 3- to 5-year-old children with CIs and peers with NH to examine the effects of device limitations on speech acquisition. If auditory deprivation alone impacts speech acquisition, children with CIs would acquire sounds on a delayed timeline but follow similar patterns of acquisition as children with NH. Alternatively, if device limitations influence speech acquisition, children with CIs would demonstrate unique patterns during development. Finally, if there is a general pattern of "Deaf speech" that results from signal degradation (whether from a CI or a hearing aid), it would be expected that productions of children with CIs and children with profound hearing impairment who use hearing aids would exhibit similar error patterns. Three main findings from this study suggest that device limitations critically impact acquisition of the /t/-/k/ contrast for children with CIs. Given the strength of this evidence from the current study, we believe that device limitations will affect speech acquisition more generally in this population.

First, children with CIs produced /k/ as accurately as /t/ prior to mastering either sound. According to normative data, children with NH acquire /t/ earlier than /k/: at age 3, /t/ is produced accurately 91% of the time compared to 77% for /k/, and at age 4, /t/ is produced with 100% accuracy while /k/ continues to lag behind (Smit et al. 1990). Thus, while children with CIs unsurprisingly acquired /t/ and /k/ later than children with NH, the acquisition of these sounds relative to each other did not follow the typical developmental trajectory. Children with CIs likely produce /t/ and /k/ with similar accuracy levels because these sounds are perceptually confusable with each other when frequency resolution is poor, and visual cues are not readily available for either sound to facilitate differentiation. Stop consonants in English also have many allophonic variations, which may make abstraction a more difficult task when the signal is both degraded and highly variable. Similar levels of accuracy across consonants for children with CIs also suggest that acquisition of /k/ was somewhat advanced or /t/ was comparatively delayed. Other research has found that children with CIs acquire /t/ later than expected (Blamey et al. 2001).

Several factors may influence acquisition of /t/ relative to /k/. First, /t/ has a diffuse spectrum with substantial energy in the high-frequency range, (past the 8000-Hz upper limit of stimulation) whereas /k/ has more energy in the mid-frequency range. The spectral peak of /k/ shifts depending on the following vowel, but typically /k/ has a more compact peak within the mid-frequency range, which could facilitate perception for children using CIs. Children with CIs also must rely on spectral energy below 8000 Hz to perceive /t/, but that energy is more likely to overlap with the spectral energy for /k/, especially in front-vowel contexts where the spectral energy

for /k/ is shifted to higher frequencies. Second, the spectrum for /t/ is relatively diffuse, with energy spread across a wide range of frequencies, much like white noise, which may be dampened by speech processing strategies designed to reduce the transmission of noise. Finally, /t/ consonants are typically shorter, and duration is more variable compared to /k/ (Umeda, 1977). Information in the transient release burst may be more difficult for children with CIs to perceive as duration decreases, so duration may be a more reliable and accessible cue for /k/. Follow-up research is planned to use electrodograms to quantify the discriminability of /t/ and /k/ in different phonological contexts (Peng et al. 2019).

The second main finding was that children with CIs produced different types of errors than children with NH, including errors not typical for children with NH during any stage of development. The errors produced by children with NH were most likely to be prevocalic voicing (largely driven by the error patterns of one child) and fronting place errors, which are well-documented as common developmental errors in children between ages 2 and 3 (McLeod, 2017; McLeod & Bleile, 2003). Manner errors—particularly affricate substitutions—are not considered developmental, and these errors were accordingly rare for children with NH in the present study. Children with CIs, on the other hand, produced manner errors more often than voicing errors. The relative infrequency of voicing errors may be because voicing contrasts are differentiated by temporal cues. Temporal envelope information is relatively preserved through CI signal processing, so temporal cues to voicing may be particularly accessible. The relative frequency of manner errors may be due to aspiration of word-initial voiceless stops: the aspiration noise following release of the

closure in stops may be perceptually similar to the frication noise following release of closure in affricates. More research is needed to determine whether affricate substitutions are also common for unaspirated stops.

Although both groups of children were equally likely to produce place errors, and the majority of place errors for both groups were "t" for /k/ substitutions, children with CIs produced more backing errors ("k" for /t/ substitutions) and stops with bilabial or glottal places of articulation, which are not typical developmental patterns. Furthermore, looking at Figures 2-4, it is clear that children using CIs were more likely to produce several different types of errors, whereas children with NH were more likely to produce only one type of error. For example, there are many children with NH who produced only place errors (circles at the top of Figure 4), and many who produced no place errors (circles at the bottom), but very few produced place errors in addition to other types of errors. The same was not true for children with CIs, the majority of whom produced more than one type of error. Error patterns produced by young children with CIs have not been studied extensively, so more work is needed to replicate this finding and extend research to other speech sounds.

The third finding is that productions of /t/ and /k/ by children with CIs in this study are different than those that have been observed in children with profound hearing impairment who use hearing aids in earlier studies (e.g., Osberger & McGarr, 1982; Smith, 1975). For example, Smith (1975) examined speech production in 8- to 15-year-old children who were Deaf or Hard of Hearing who used hearing aids and attended oral schools. The children in Smith (1975) produced both /t/ and /k/ less accurately (37% and 40%, respectively) than the younger children in the current study

who used CIs (70% for /t/ and 65% for /k/). Error patterns also differed. The majority of errors in Smith (1975) were omissions (51% for /t/ and 46% for /k/). We cannot directly compare the percentage of omissions in the current study to earlier studies because in our coding system, [Other] included omissions as well as distortions, fricative substitutions, approximant substitutions, glide substitutions, and other articulations that were not easily transcribed. Nevertheless, the percentage of [Other] errors produced by children with CIs in the current study was relatively low: 12% for /t/ and 11% for /k/. Thus, even if every production transcribed as [Other] was an omission, these errors would account for less than 25% of the errors produced by children with CIs. Furthermore, the most common substitutions for /t/ and /k/ in these earlier studies were glottal stops. By contrast, in the current study, 76% of place-of-articulation errors made by children with CIs were either "t" for /k/ or "k" for /t/ substitutions, and glottal stops were relatively rare. Also, manner substitutions (e.g., fricatives or affricates instead of stops) were common, especially for target /t/, which are rare substitutions for stop consonants, even for children with hearing impairment (Obserger & McGarr, 1982). The differences in error patterns between the two groups of children with profound hearing impairment suggest that device limitations uniquely impact speech acquisition for children using CIs. If auditory deprivation or a more general response to signal degradation (whether through a hearing aid or a CI) was the explanation, we would expect similar error patterns.

It is of interest to note that only one comparison (Model 3c, for /k/ in front-vowel contexts) showed a significant effect of Group on centroid frequency (the acoustic measure that differentiates /t/ from /k/). This finding could be considered

unsurprising, because the acoustic analysis was performed only on correct productions. However, both Reidy et al. (2017) and Todd et al. (2011) found that correct productions of /s/ by children with CIs were spectrally distinct from correct productions of /s/ by children with NH. The difference across studies may be related to greater motor demands required to produce sibilant fricatives relative to stops. It is also possible that acoustic differences exist across groups for /t/ and /k/ productions, but the differences are not fully captured by centroid frequency.

There were some significant interactions between vowel context and the measures of accuracy and spectral features. These effects were different for children with CIs compared to children with NH, which may be an effect of spectral degradation. For children with NH, vowel context had no effect on accuracy. Children with CIs, on the other hand, produced /k/ less accurately in front-vowel contexts compared to back-vowel contexts. Both groups produced /k/ in front-vowel contexts with higher centroid frequency than /k/ in back-vowel contexts (the expected effect of coarticulation), but children with NH produced /k/ in front-vowel contexts with significantly higher centroid than children with CIs, suggesting they coarticulate more than children with CIs. In fact, children with NH produced /t/ and /k/ in front-vowel contexts with equal centroid frequencies, providing evidence of spectral overlap between these sounds in front-vowel contexts, which could exacerbate perceptual confusability. However, it should be noted that children with NH produced more closer-to-correct intermediate productions (n = 51) than children with CIs (n = 46), and most of these were [k:t] (for target /k/) in front-vowel contexts (n = 31). It is possible that these intermediate productions from children with NH contributed to the

spectral overlap of /t/ and /k/ observed in front-vowel contexts. It is also possible that children with CIs do indeed, coarticulate less than children with NH, as the cues to coarticulation are primarily spectral and we know that coarticulation is at least partially learned and planned, rather than reflecting language-universal motor implementation (Cychosz, Munson, & Edwards, 2021; Noiray et al., 2019). Future research could examine whether children with CIs do indeed produce /k/ in front-vowel contexts with a more back tongue position or if they produce front vowels differently than children with NH.

The findings related to context effects are difficult to interpret in terms of device limitations, because studies examining the effects of vowel context on articulation do not exist for younger children with NH. It is possible that vowel context affects accuracy at some point during development, and the degree of coarticulation increases as articulatory skills are refined. Additional research is necessary to determine whether the effects of vowel context observed in the current study are unique to children with CIs, or if these patterns are also found during earlier stages of development for children with NH. It may also be worthwhile in future research to measure each vowel's spectral features to determine whether there are sub-phonemic differences in vowel productions across groups.

Although data on younger children with NH are available, one limitation of the current study was not having a direct comparison group of hearing-age matched peers. This was difficult due to the number of children with CIs who received implants by their first birthdays and were tested in the laboratory at age 3. Children with NH who had the same amount of hearing experience as our children with CIs

would have been 2 years old, which would have reduced the number of familiar words available for the experiment. Also, completing the task with the number of repetitions necessary for analyses would not have been feasible with children that young.

There are some challenges to separating the effects of auditory deprivation and device limitations. It is possible that auditory deprivation in the first year of life alters brain development in ways that impact later speech acquisition, and this early disruption could manifest as qualitative differences rather than straightforward delays. Furthermore, development of neuromotor control for speech may be impacted by auditory deprivation in the first year of life by ineffective perceptual-motor feedback loops. Although gross and fine motor skills develop in the absence of auditory input, it is possible that specialized control for speech requires access to reliable perceptual-motor feedback loops in the first year of life. Extensive research is needed to compare speech acquisition in young children with CIs who received auditory input in the first year of life and those who did not in order to fully comprehend the relative effects of auditory deprivation and device limitations.

Conclusion

The current study investigated how speech acquisition is affected by the unique perceptual constraints experienced by young children using CIs. Patterns observed in the present study are not sufficiently characterized by delayed acquisition alone. The current findings suggest that speech development is also altered by device limitations. Children with CIs acquired /t/ and /k/ consonants following a different developmental trajectory than children with NH. Children with CIs also produced

different types of errors than children with NH, and their errors did not reflect typical developmental patterns. Finally, the effects of vowel context were not similar across groups: vowel context had a larger effect on accuracy but a smaller effect on coarticulation for children with CIs compared to children with NH.

Knowledge of how device limitations specifically impact perception, learning, and production of speech is critical for developing more effective approaches to speech therapy and improving speech processing strategies. Speech-language pathologists should consider the perceptual constraints and confusability of consonants when assessing and treating children who use CIs. It may be useful to target manner of articulation before targeting place contrasts. Speech-language pathologists should also provide natural models of speech targets, because acoustic features may be transmitted differently when emphasized. For example, emphasizing articulation of a word-initial /t/ may increase perceptual confusability with the affricate "ch," because the noise of aspiration may become more similar to the noise of frication following the release of closure.

Future work inspired by the current study includes assessing listeners' perception of /t/ and /k/ productions at both the token- and word-levels, as well as analyzing accuracy and error patterns of several later-acquired sounds and contrasts likely to be affected to varying degrees by perception and production constraints (e.g., the /s/-/ʃ/ and /ɹ/-/w/ contrasts, consonant clusters).

Chapter 4: Effects of device limitations on acquisition of the /s/-/ʃ/ contrast in children with cochlear implants

Abstract

The primary purpose of this paper was to investigate the effects of reduced spectral resolution in the auditory signal on acquisition of the spectral /s/-/ʃ/ contrast in 3- to 5-year-old children who use cochlear implants (CIs). We hypothesized that children with CIs would not demonstrate straightforward delays relative to peers with NH. Instead, they would demonstrate unique patterns of speech development that are more related to signal degradation than a period of auditory deprivation. Although acquisition of the /s/-/ʃ/ contrast has been studied previously in older children with CIs, the current study extended this research to younger children, so that patterns of production could be compared across groups of children who were still in the process of acquiring this contrast. A secondary purpose of this paper was to examine the robustness of contrast for both /t/ vs. /k/ and /s/ vs. /ʃ/ in adults with NH, children with NH, and children with CIs.

Transcription analyses were used to compare overall accuracy and error patterns of word-initial /s/ and /ʃ/ produced by the same groups of children described in Chapter 3 (Johnson et al., 2021). Centroid frequencies of correct productions were also calculated and compared across groups. Robustness of contrast was quantified using the method described in Chapter 2 for /t/ and /k/, and robustness of both the /s/-/ʃ/ and /t/-/k/ contrasts are compared.

Children with CIs produced /s/ significantly less accurately than their peers with normal hearing who were matched in terms of age, sex, and maternal education,

but they produced /ʃ/ with equal accuracy. Acoustic analyses revealed that children with CIs also produced /s/ with lower centroid frequencies than children with NH, even though there was no difference across groups for /ʃ/ centroids. Although the errors produced by both groups were equally likely to be place-errors or manner-errors, children with CIs produced more deletions and distortions compared to their peers. Adults produced more robust contrasts for both sets of consonants relative to children. Also, children with NH produced more robust /s/-/ʃ/ contrasts compared to children with CIs, but there was no difference across groups for /t/-/k/ contrasts. These idiosyncratic patterns of production observed for the children with CIs provide evidence that the process of learning phonetic categories from a spectrally degraded speech signal is fundamentally different, rather than simply delayed.

Introduction

Acquisition of the /s/-/ʃ/ Contrast

As discussed in Chapters 1 and 3, cochlear implants are one of the most successful technological innovations in modern history. However, even children who receive their implants at the earliest possible age demonstrate delays in speech acquisition compared to their peers with normal hearing after many years of experience with the device (e.g., Spencer & Guo, 2013). There are two possible, non-mutually exclusive, explanations for slower development and overall worse speech outcomes in children with CIs compared to their peers with NH. First, children who are born with profound hearing loss experience a period of auditory deprivation prior

to implantation, typically at least one year. Second, the signal delivered by a CI is highly degraded (for additional details, refer to Chapter 3).

Because spectral information is particularly distorted—or eliminated completely in the case of frequencies above 8,000Hz—when speech is processed by CIs, learning to produce speech-sound contrasts that are differentiated auditorily by spectral cues is more likely to be impacted compared to contrasts differentiated by temporal or amplitude information. I hypothesized that children with CIs would demonstrate not only delayed acquisition, but also idiosyncratic patterns of acquisition compared to their peers with NH while learning place-of-articulation contrasts, which are cued spectrally in English. The results presented in Chapter 3 (Johnson et al., 2021) for the early-acquired /t/-/k/ contrast supported this hypothesis: 3- to 5-year-old children with CIs produced both /k/ and /t/ less accurately than their peers with NH, but they also produced /k/ and /t/ with *equal* accuracy. According to normative data, children with NH typically acquire /t/ before /k/ (Smit et al., 1990; Crowe & McLeod, 2021), so for the children with CIs in Chapter 3, the order of acquisition during development of this contrast was atypical. Furthermore, children with CIs produced different types of errors compared to their peers with NH (and normative data for younger children with NH). The most common errors for children with CIs were manner errors, whereas the most common errors for children with NH are voicing errors (Johnson et al., 2021; McLeod, 2017; McLeod & Bleile, 2003). Consistent with the idea that auditory deprivation leads to delayed acquisition, children with CIs did demonstrate a delay in acquiring the /t/-/k/ contrast. However, the atypical patterns that were observed for both the order of acquisition and the types

of errors produced are more consistent with the explanation that signal degradation impacts speech development.

The /t/-/k/ contrast studied in Chapter 3 is an early-acquired contrast: most children produce both /t/ and /k/ accurately by age 3-4 (Crowe & McLeod, 2021; Smit et al., 1990). Consistent with normative data, most of the children with NH analyzed in Chapter 3 had already acquired /t/ and /k/, evidenced by producing both sounds with ceiling-levels of accuracy (above 90% for both consonants across vowel contexts). This chapter will focus on production of the sibilant fricatives, /s/ and /ʃ/. Development of these fricatives is protracted, and they are not typically mastered until age 5 (Smit et al., 1990; Crowe & McCloud, 2021). By studying /s/ and /ʃ/ produced by 3- to 5-year-old children, it will be possible to compare patterns of accuracy, error types, and spectral features across groups of children who are all in the process of acquiring this contrast.

The protracted development of /s/ and /ʃ/ is most likely due to increased motor demands required to produce these fricatives. While stop consonants are produced with ballistic tongue movements, fricatives require precise tongue placement and contouring to control airflow (Kent, 1994). The English sibilant fricatives are produced by the tongue creating a narrow constriction at the alveolar place (for /s/) and at the alveopalatal place (for /ʃ/). A noise source is generated when air flows through this narrow constriction and becomes turbulent. Energy in the spectra varies as a function of the size of the resonant cavity anterior to the constriction (in addition to other obstructions, such as teeth). For any given speaker or phonological context, /ʃ/ is formed farther back in the mouth than /s/. The result is a relatively large

resonant cavity anterior to the constriction, which corresponds to energy concentrated at lower frequencies for /ʃ/ compared to /s/. This precise tongue placement is challenging motorically, even for children with NH. Children likely use an auditory-articulatory feedback loop to master these sounds: as children adjust tongue position, they listen for changes in the acoustic signal, and continue adjusting until their productions sound similar to the models in their environment.

The acquisition of the /s/ vs. /ʃ/ contrast is even more protracted for children who use cochlear implants (Faes & Gillis, 2016; Serry & Blamey, 1999; Blamey et al., 2001; Warner-Czyz & Davis, 2008). The ability to learn to produce a contrast is at least somewhat related to the ability to perceive that contrast (e.g., Matthies et al., 1994; Stelmachowitz et al., 2004). Thus, children with cochlear implants take a longer time and may require explicit training to learning the /s/-/ʃ/ contrast due to the factors described in Chapters 1 and 3. Numerous studies comparing /s/ and /ʃ/ productions have consistently found that children with CIs produce /s/ with lower spectral centroid (or peak) frequency than peers with NH (Liker, Mildner, & Sindija, 2007; Neumeyer, Schiel, & Hoole, 2015; Uchanski & Geers, 2003). Several studies that used a similar methodology to the current studies (i.e., eliciting multiple productions of both /s/ and /ʃ/ in word-initial position using a picture-prompted, auditory repetition task) found that even though spectral features for /ʃ/ across groups were similar, lower centroid frequencies for /s/ resulted in a less robust /s/-/ʃ/ contrast for children with CIs (Reidy et al., 2017; Todd et al., 2011). Reidy and colleagues (2017) also showed that, even within correct productions, these acoustic differences had perceptual consequences: adults judged /s/-initial words produced by children

with CIs as less intelligible than /ʃ/-initial words, but they perceived the /s/- and /ʃ/-initial words produced by children with NH as equally intelligible.

The Reidy et al. (2017) and Todd et al. (2011) studies focused on 4- to 7-year-old and 4- to 10-year-old children with CIs and age-matched peers with NH. Thus, many of the children with NH had already learned to produce the /s/-/ʃ/ contrast, as was the case in Chapter 2 for the /t/ vs. /k/ contrast. The purpose of this study was to compare productions of /s/ and /ʃ/ across 3- to 5-year-old children with and without CIs in order to examine patterns that occur while children are still in the process of acquiring these sounds.

As in Chapter 3, this study investigated whether there were differences in accuracy, error patterns, and spectral features for /s/ and /ʃ/ produced by children with CIs compared to their peers with NH. Based on the work of Reidy et al. (2017) and Todd et al. (2011), I hypothesized an interaction where children with CIs would produce both sounds less accurately than their peers, and both groups of children would produce /ʃ/ more accurately than /s/, but the difference in accuracy across consonants would be larger for children with CIs. In terms of errors, I hypothesized that both groups of children would produce a greater number of manner and place errors than voicing errors, because voicing errors are not typically observed for target fricatives during development, and voiced fricatives contain more energy at lower frequencies, making them unlikely to be auditorily confused with voiceless fricatives when processed by CIs. Within manner errors, however, I hypothesized that children with CIs would produce more deletions and distortions than their peers with NH. Finally, I hypothesized that children with CIs would produce /s/ with lower centroid

frequency than children with NH, thereby reducing their robustness of contrast. If true, these results would extend the evidence from Chapter 3 to a late-acquired contrast, showing that signal degradation impacts acquisition of spectral contrasts for children with CIs.

The acoustic analysis presented in Chapter 3 and described above for /s/ and /ʃ/ focuses on production of individual sounds. But children do not simply learn individual sounds when they acquire a phonology—they learn a system of contrasts. In fact, children sometimes produce a systematic contrast between two sounds before they produce one of the sounds correctly. In this case, children are producing *covert contrasts*, where two different target phonemes are perceived and transcribed as belonging to the same phonemic category despite a measurable and systematic phonetic distinction (e.g., Li, Edwards, & Beckman, 2009; Macken & Barton, 1980; Maxwell & Weismer, 1982; Munson et al., 2010). Covert contrasts are clinically significant: children being treated for speech-sound disorders who produce a covert contrast progress faster through therapy than children who produce no contrast (Tyler, Figurski, & Langsdale, 1993). One explanation for this finding is that children who produce any contrast– even a subphonemic one—are demonstrating a foundational knowledge upon which articulation therapy can build, whereas children who produce no contrast must first learn to perceive the contrast.

A secondary purpose of this paper was to quantify robustness of both the /t/-/k/ and /s/-/ʃ/ contrasts and compare robustness of these contrasts across adults and children with and without CIs. As reviewed in Chapter 2, Holliday et al. (2015) proposed using the percentage of tokens correctly classified by a model based on a

single acoustic measure as a measure of robustness of contrast. For both the /t/-/k/ and /s/-/ʃ/ contrasts, I predicted that adults would have more robust contrasts than children, and children with NH would have more robust contrasts than children with CIs.

This paper differs from previous research in two important ways. First, it examines acquisition of the /s/-/ʃ/ contrast in 3- to 5-year-old children with and without CIs, a younger age than previous research. Second, it examines the robustness of two contrasts (/t/-/k/ and /s/-/ʃ/) across three groups (adults with NH, children with NH, and children with CIs).

Materials and Methods

Participants

A subset of 16 out of the 21 adults described in Chapter 2 also produced /s/ and /ʃ/ tokens, which will be analyzed in this chapter.

The same 26 children (and their matches) described in Chapter 3 also produced /s/ and /ʃ/ tokens, which will be analyzed in this chapter. However, longitudinal data were not transcribed for four of the children who had longitudinal /t/ and /k/ data. For this /s/-/ʃ/ analysis, 19 children from each group participated in 1 session, 4 children participated twice, and 3 children participated 3 times, yielding productions from 52 different children over a total of 72 sessions.

Table 1. Participant Demographic Table.

Group	Age	Boys Girls	*EVT-2*[a] Standard score (*SD*): 100 (*15*)	*PPVT-4* [b] Standard score (*SD*): 100 (*15*)	*GFTA-2*[c] Standard Score (*SD*): 100 (*15*)
CI n = 26	50 months range: 31-65 n = 36	16 20	98 (*20*) range: 46 – 131 n = 35	94 (*22*) range: 40 – 139 n = 35	74 (*21*) range: 39 – 107 n = 32
NH n =26	50 months range: 32-66 n = 36	16 20	117 (12) range: 88 – 134 n = 36	120 (11) range: 94 – 140 n = 20	91 (12) range: 67 – 113 n = 24

[a] *Expressive Vocabulary Test-2nd Edition* (Williams, 2007)
[b] *Peabody Picture Vocabulary Test-4th Edition* (Dunn & Dunn, 2007)
[c] *Goldman-Fristoe Test of Articulation-2nd Edition* (Goldman & Fristoe, 2000)

Stimuli

Auditory prompts paired with relevant picture cues comprised stimuli for the experiment. In total, there were 16 different familiar words with initial /s/ targets and 16 with initial /ʃ/ targets. The process of selecting words and distributing them across separate wordlists was identical to that of choosing /t/- and /k/-initial words. To reiterate, words were selected if 90% of children within our targeted age groups understand and produce them according to databases with published age-of acquisition norms (Dunn & Dunn, 2007; Fenson et al., 2007; Morrison, Chappell, & Ellis, 1997). Words used to elicit target responses were distributed across three wordlists, each with a unique set of words designed to be familiar for children within a specific age range. The target words and the number of times they appeared in each wordlist are presented in Appendix E. As children got older, words with older ages of

acquisition were added to each wordlist, thus increasing the number, variety, and complexity of the words. Children were assigned to wordlists based on age at test as well as prior participation; for example, a 44-month-old child participating for the first time would be assigned the first wordlist, but a 44-month-old child participating for the second time would be assigned the second wordlist (because they had previously completed the first wordlist at 32 months of age).

The first wordlist was administered to 8 children with NH (ages 32-39 months) and 13 children with CIs (ages 31-44 months). Wordlist 1 included 5 different /ʃ/-initial words and 8 different /s/-initial words, each repeated 2-4 times to elicit a total of 16 /ʃ/ consonants and 16 /s/ consonants (equally balanced across vowel contexts).

The second wordlist was administered to 12 children with NH (ages 43-52 months) and 18 children with CIs (43-65 months). Wordlist 2 included 10 different /ʃ/-initial words and 14 different /s/-initial words, each repeated 1-2 times to elicit a total of 14 /ʃ/ consonants (8 in front-vowel contexts) and 16 /s/ consonants (9 in front-vowel contexts).

The third wordlist was administered to 16 children with NH (ages 51-66 months) and 5 children with CIs (ages 55-66 months). Wordlist 3 included 10 different /ʃ/-initial words and 12 different /s/-initial words, each repeated 1-2 times to elicit 16 /ʃ/ consonants (8 in each vowel context) and 16 /s/ consonants (9 in front-vowel contexts).

Each adult completed both Wordlist 2 and Wordlist 3 (on separate testing days).

Procedure

The procedure used to elicit /s/ and /ʃ/ tokens was identical to that described in Chapter 2 and 3 for the /t/ and /k/ tokens.

Coding

Transcribing and event-marking were completed in *Praat* (Boersma & Weenink, 2018). First, research assistants trained in phonetics (including the author, undergraduate, and graduate students) listened to each recording session in its entirety, marking word-onset and word-offset boundaries around each target response. Sometimes children produced multiple target responses; in these cases, boundaries for each response were marked, and the context of each response was noted (e.g., clinician prompted the child to repeat the word, child self-corrected an error). Next, phoneticians transcribed the manner, place, and voicing of the initial consonant of each target response. Phoneticians transcribed the child's immediate response to the auditory prompt—even if the utterance contained an articulation error—unless any of the following exclusionary criteria were met: 1) the child's response was obscured by background noise, such that some or all of the target consonant was neither clear enough to transcribe nor sensible to analyze acoustically; 2) child had a false start or self-corrected an error; 3) child's response was significantly louder than during calibration and caused the recording to clip. If an initial response was excluded based on the above criteria, the child's second response

was transcribed (if available); trials that yielded no transcribable responses were treated as missing data and excluded from analyses. Transcribers also flagged trials where the child's response overlapped with the stimulus. Trials with overlapping responses were typically audible, and thus transcribable; however, due to interference, they were excluded from spectral analyses.

After determining which response to transcribe, phoneticians judged the manner and place of articulation of the target consonant. (Refer to Chapters 2 and 3 for the procedure used to transcribe and event-mark target /t/ and /k/ consonants.)

For /s/ and /ʃ/ targets, options for manner transcriptions included [Sibilant fricative], [Non-sibilant fricative] [Sibilant affricate], [Stop], and [Other]. The [Other] category for manner included deletions, distortions, glide substitutions, approximant substitutions, and other articulations that were not easily categorized. For productions transcribed as Sibilant fricatives, options for place transcription included [s], [S], [s:S] (intermediate, closer to [s]), [S:s] (intermediate closer to [S]), and [other]. The [other] category for place was used when the child produced a consonant sequence (e.g., [sS]). The [other] category was also designed to include voicing errors (e.g., [z]), but not a single child produced a fully-voiced sibilant fricative substitution.

For productions transcribed as [Sibilant fricative], coders marked the locations on the waveform corresponding to the onset of turbulence and the onset of voicing. The onset of turbulence was defined as the beginning of aperiodic, high-frequency frication noise visible in the waveform with white noise in a frequency band above 1000Hz. The onset of voicing was defined as the first upward swing from the zero-crossing followed by a stable, quasi-periodic pattern of voicing. In some cases, the

child produced aspiration between the fricative and the vowel; in these cases, the offset of frication was also marked.

For /s/ and /ʃ/ targets, productions were scored as Correct (i.e., Accuracy = 1) when they were transcribed as [Sibilant fricative] with a place transcription of [s] or [s:S] for target /s/ or [S] or [S:s] for target /ʃ/. Incorrect productions were coded as Manner errors (i.e., Manner error = 1) when the manner of articulation was coded as anything other than [Sibilant fricative] or [Non-sibilant fricative; Place errors (i.e., Place error = 1) when the manner of articulation was coded as [Sibilant fricative], but the place of articulation did not match the target consonant (e.g., "sh" for /s/); or Voicing errors (i.e., Voicing error = 1) when the manner of articulation was [Sibilant fricative], but the time between the onset of turbulence and the onset of voicing was <40 ms or if the onset of voicing was marked prior to the offset of turbulence.

A random 10% of files (7 files total, 1 from child with NH, 6 from children with CIs) were coded by two researchers for reliability purposes. An additional four files from children with NH who were not included in this study but were transcribed by both researchers for the larger project were also included for this reliability calculation. Agreement between the two transcribers on whether the sound was produced accurately was 99% for the children with CIs (one disagreement out of 186 tokens transcribed) and 94% for children with NH (7 disagreements out of 124 tokens transcribed).

Only sessions with data used in the acoustic analyses were included in the following reliability calculations. There were 111 tokens for which both transcribers agreed acoustic analyses were appropriate (i.e., the child produced the target correctly

and the response was not obscured by background noise). Of these, transcribers agreed on which response to transcribe in 92% cases. Root mean square (RMS) values were calculated to quantify the differences between transcribers in the locations of fricative midpoints and the onsets of voicing, as well as the resulting centroid frequency calculation. The RMS for fricative midpoint was 3.9ms. The RMS for voicing onset was 3.1ms. The RMS for centroid frequency was 31Hz.

Spectral Measures

(Refer to Chapters 2 and 3 for the procedure used to calculate centroid frequency for the /t/ and /k/ tokens.)

For the English voiceless sibilant fricatives /s/ and /ʃ/, information from noise spectra has been particularly useful for differentiating among places of articulation (Behrens & Blumstein, 1988; Forrest et al., 1988; Hughes & Halle, 1956; Heinz & Stevens, 1961; Jongman, Wayland, & Wong, 2000; Nittrouer et al., 1989; Nittrouer 1995; Strevens, 1960). These fricatives are produced by the tongue creating a narrow constriction at the alveolar place (for /s/) and at the alveopalatal place (for /ʃ/). The noise source is generated when air flows through this narrow constriction and becomes turbulent. Energy in the spectra varies as a function of the size of the resonant cavity anterior to the constriction (in addition to other obstructions, such as teeth). For any given speaker or phonological context, /ʃ/ is formed farther back in the mouth than /s/. The result is a relatively large resonant cavity anterior to the constriction, which corresponds to energy concentrated at lower frequencies for /ʃ/ compared to /s/. Studies analyzing /s/ and /ʃ/ spectra have confirmed that /s/ has

prominent spectral peaks at relatively higher frequencies than /ʃ/ (Behrens & Blumstein, 1988; Hughes & Halle, 1956; Heinz & Stevens, 1961; Strevens, 1960). In addition to the location of spectral peaks, analyses of the first spectral moment (i.e., the mean, or *centroid* frequency) have also confirmed that /s/ has a higher centroid frequency than /ʃ/ (Forrest et al., 1988; Jongman, Wayland, & Wong, 2000; Newman et al., 2001; Nittrouer et al., 1989; Nittrouer 1995). Thus, centroid frequency is an excellent acoustic measure to differentiate between /s/ and /ʃ/. Furthermore, unlike in the case of the /t/-/k/ contrast, it is relatively unaffected by vowel context.

For /s/ and /ʃ/ tokens coded as correct and unobstructed by noise, acoustic spectra were estimated, and centroid frequencies were calculated using custom scripts in the *R* programming environment (R Core Team, 2013). Spectral measures were computed following the procedure described by Holliday et al. (2015). To summarize: a 40-ms analysis window with zero-padding was centered around the temporal midpoint of the fricative. Within the analysis window, the acoustic spectrum of the waveform was estimated using an 8^{th}-order multitaper spectrum ($K = 8$, $NW = 4$). These acoustic spectra were then transformed into normalized power spectra with cutoff frequencies of 300 Hz and 20,000 Hz. The normalized power spectra were treated as a probability mass function over frequency, and the centroid frequency for each token was the mean value of this distribution. This procedure was identical to that used for /t/ and /k/ tokens in Chapters 2 and 3 with three exceptions: the size of the window (25 ms for stops, 40 ms for fricatives), the location of the window (at the release burst for stops, centered around the midpoint for fricatives), and the range of

values within which the spectra were normalized (926-10,000 Hz for stops, 300-20,000 Hz for fricatives).

Accuracy Analysis

An identical approach was used to analyze accuracy of /s/ and /ʃ/ targets as described in Chapter 3 for /t/ and /k/ targets (i.e., four homologous logistic mixed-effects regression models with different reference categories were used to compare accuracy within and across groups, consonants, and vowel contexts). To account for multiple models, an adjusted alpha level of $p = 0.0125$ was used.

Error Pattern Analysis

The approach to analyzing errors was slightly different than the approach used for /t/ and /k/ targets described in Chapter 3 for two reasons. First, none of the children produced voicing errors, so the only types of errors that could be analyzed were manner and place errors. Second, because place-of-articulation was only transcribed for sibilant fricatives, these two types of errors were mutually exclusive. Therefore, only two homologous models (with the structure described in Chapter 3, Equation 2) were necessary to make all comparisons. These models predicted the log-likelihood that an incorrect production contained a manner error (or a place error) based on fixed effects of Group, Target Consonant, and the 2-way interaction between Group and Target Consonant, as well as the participant-level random

intercept. To account for multiple models, an adjusted alpha level of $p = 0.025$ was used.

Spectral Feature Analysis

An identical approach was used to analyze centroid frequencies of /s/ and /ʃ/ targets as described in Chapter 3. To account for multiple models, an adjusted alpha level of $p = 0.0125$ was used.

Calculating Robustness of Contrast

An identical approach was used to calculate robustness of contrast as described in Chapter 2 for /t/ and /k/, following the procedure first described by Holliday et al. (2015) with the addition of vowel context in the model used make predictions. To review: for tokens that were produced correctly based on transcription, a mixed-effects logistic regression model was fit that predicted the log-likelihood that a given token was either /s/ or /ʃ/ (or /t/ or /k/) based on fixed effects of intercept, centroid frequency, vowel context, and the interaction between centroid and vowel context, as well as speaker-level random intercepts and random slopes for centroid frequency (see Equation 1). If the model classified the token correctly, *Predicted Accuracy* for that token equaled 1. Then, within each speaker, robustness of contrast was calculated as the percentage of their tokens correctly classified by the model.

Robustness of contrast was calculated for adults' and children's /t/-/k/ and /s/-/ʃ/ contrasts. The sixteen adults with two sessions each were included in the robustness analysis (the five adults with /t-/k/ data but no /s/-/ʃ/ data were excluded). Previous research has suggested that within-speaker variability can be assessed with a sample of four tokens. (Newman et al., 2001; Uchanski et al., 1992). Thus, sessions from children (and their matches) who did not produce at least two correct tokens of each consonant per vowel context were excluded. After excluding these sessions with too few analyzable tokens, the final analyses included 24 sessions (12 from each group) for the fricative contrast and 50 sessions (25 from each group) for the stop contrast.

To compare robustness of the adults' /t/-/k/ contrasts to their /s/-/ʃ/ contrasts, a new data frame was created that combined all of the adults' data and added a new a variable for "Contrast" (with two levels, "/t/-/k/" and "/s/-/ʃ/"). Then, a mixed-effects logistic regression model was used to predict the log-likelihood that a prediction was accurate based on fixed effects of intercept, contrast, and speaker-level random intercepts.

To compare robustness of contrast across children and adults, two new data frames were created. One combined the adults' and children's /t/-/k/ data, and the other combined the adults' and children's /s/-/ʃ/ data. Two models were used (one for each contrast) to predict the likelihood that a prediction was accurate based on fixed effects of intercept, group (levels = adults, children with NH, and children with CIs), and speaker-level random intercepts.

Finally, to compare robustness of contrast across children with and without CIs, a new data frame was created that combined all of the children's data and added the "Contrast" variable. Two homologous models with different reference categories were fit to examine the relationships between all levels of predictors. These models predicted the log-likelihood that a prediction was accurate based on fixed effects of intercept, Group (CI or NH), Contrast (/t/-/k/ or /s/-/ʃ/-), the interaction between Group and Contrast, and the speaker-level random intercepts. To account for multiple models, an adjusted alpha level of 0.025 was used to determine significance.

Results[7]

Accuracy Results

The children in this study produced 2206 responses (1130 with target /s/, 1076 with target /ʃ/). Children with CIs produced target /s/ (557 tokens) with 48% accuracy (*SD* = 50%) and target /ʃ/ (524 tokens) with 58% accuracy (*SD* = 49%). Children with NH produced target /s/ with 74% accuracy (*SD* = 44%) and target /ʃ/ with 65% accuracy (*SD* = 48%). Table 2 shows the percentage of tokens produced correctly by group, target consonant, and vowel context.

[7] Detailed model results for all analyses included in this chapter are included in Appendix D.

Table 2. The mean percentage (and *SD)* of /s/ and /ʃ/ tokens produced correctly by children with CIs and children with NH in front- and back-vowel contexts.

	/s/		/ʃ/	
	Front-vowel context	**Back-vowel context**	**Front-vowel context**	**Back-vowel context**
CI	47% (50%)	48% (50%)	59% (49%)	58% (50%)
NH	75% (44%)	73% (44%)	63% (48%)	66% (47%)

Significant differences between the groups in accuracy are illustrated in Figure 1. The results of Models 1a-1d are presented in Appendix D.

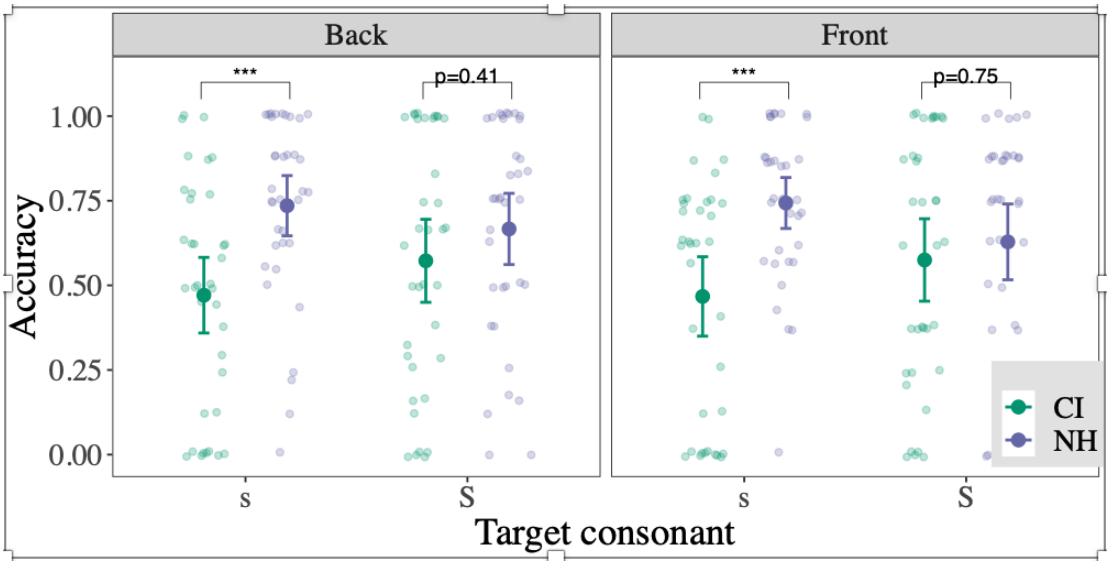


Figure 1. Mean accuracy scores of /s/ and /ʃ/ productions for each group (large circles) with ±2 standard error bars and individual data (smaller transparent circles) for word-initial /k/ and /t/ tokens in both back- and front-vowel contexts. *** indicates $p<0.001$.

Model 1a (reference: CI, target /ʃ/, back-vowel context) showed no significant effects of any variable or interaction after the adjusted alpha-level was applied. The effect of Target Consonant and the interaction between Target Consonant and Group

showed non-significant but noteworthy trends, indicating that children with CIs produced /s/ in back-vowel contexts less accurately than /ʃ/ in back-vowel contexts ($\hat{\beta}_2$ = -0.86, *SE* = 0.36, *z* = -2.38, *p* = 0.018), but children with NH produced /s/ more accurately than /ʃ/ ($\hat{\beta}_4$ = -1.20, *SE* = 0.50, *z* = 2.37, *p* = 0.018).

Model 1b (reference: CI, /s/, front-vowel context) showed significant main effects of Group ($\hat{\beta}_1$ = 1.80, *SE* = 0.47, *z* = 3.86, *p* < 0.001) and Target Consonant ($\hat{\beta}_2$ = 0.95, *SE* = 0.36, *z* = 32.65, *p* < 0.01), as well as a significant interaction between Group and Target Consonant ($\hat{\beta}_4$ = -1.62, *SE* = 0.50, *z* = -3.23, *p* < 0.01). The main effect of Group indicates that children with NH produced /s/ in front-vowel contexts more accurately than children with CIs. The main effect of Target Consonant indicates that children with CIs produced /ʃ/ in front-vowel contexts more accurately than /s/ in front-vowel contexts. The interaction indicates that in front-vowel contexts, the difference in accuracy between groups was larger for /s/ compared to /ʃ/ productions.

Model 1c (reference: NH, /ʃ/, front-vowel context) showed only a significant interaction between Group and Target Consonant ($\hat{\beta}_4$ = -1.62, *SE* = 0.50, *z* = -3.23, *p* < 0.01). This interaction indicates the groups produced /ʃ/–but not /s/–in front-vowel contexts with statistically equivalent accuracy levels.

Model 1d (reference: NH, /s/, back-vowel context) showed a significant main effect of Group ($\hat{\beta}_1$ = -1.67, *SE* = 0.46, *z* = -3.62, *p* < 0.001). The main effect of Group suggests that children with CIs produced /s/ in back-vowel contexts less accurately than children with NH.

Taken together, these results indicate that children with CIs produced /ʃ/ as accurately as children with NH in both front- and back-vowel contexts, but they produced /s/ significantly less accurately across vowel contexts. These results also show that in front-vowel contexts, children with NH produced /s/ and /ʃ/ with equal accuracy (and a trend towards higher accuracy for /s/ than /ʃ/) but children with CIs produced /ʃ/ significantly more accurately than /s/.

Error Pattern Results

Out of the 2206 tokens transcribed, 854 were produced incorrectly (509 errors by children with CIs, 345 by children with NH) and included in this analysis of error patterns. Children with CIs produced more errors than children with NH, and both groups produced more manner errors than place errors. The total number of each type of error across groups and target consonants, and the percentage of each type of error within each group, are presented in Table 3, and the proportion of incorrect productions that included manner errors is illustrated in Figure 2.

Table 3. The number (and percentage) of incorrectly produced /s/ and /ʃ/ targets containing Manner and Place errors for children with CIs and children with NH.

Group	Target /s/	Target /ʃ/	No. of Tokens (% of Total Errors*)
Manner Errors			
CI	201	166	367 (72%)
NH	99	116	215 (62%)
Place Errors			
CI	90	52	142 (28%)
NH	50	80	130 (38%)

Manner errors included productions transcribed as Non-sibilant plosives, Sibilant affricates, or Other. Place errors included productions transcribed as Non-sibilant fricatives, or productions transcribed as sibilant fricatives where the transcribed consonant did not match the target consonant.

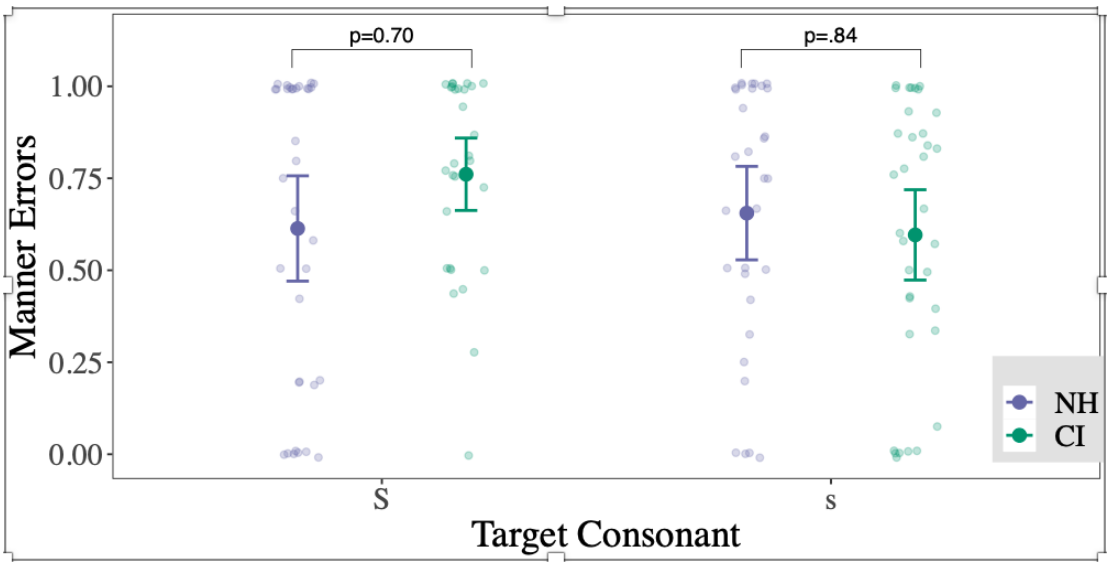


Figure. 2. Mean proportion of incorrect /s/ and /ʃ/ productions that contained Manner errors for each group (large circles) with ±2 standard error bars and individual data (small circles).

Predicting the log-likelihood that an incorrect production contained a Manner error, neither model 2a (reference: CI, target /s/) nor model 2b showed any significant effects. The results of these models indicate that children with CIs and children with NH were equally likely to produce either a manner error or a place error, and these two types of errors were equally distributed across target consonants.

Spectral Feature Results

The children in this study produced 1352 target consonants correctly. After excluding 160 productions that overlapped with the stimulus (and one production where the time between turbulence onset and voicing onset was less than 40ms), there were 1192 analyzable tokens: 264 /ʃ/ and 230 /s/ tokens produced correctly by children with CIs, and 332 /ʃ/ and 366 /s/ tokens produced correctly by children with NH. Mean centroid frequencies for each group across target consonants and vowel

contexts is presented in Table 4. Group differences in spectral features are illustrated in Figure 3.

Table 4. Centroid values (and *SD*) of /s/ and /ʃ/ tokens produced correctly by children with CIs and children with NH in front- and back-vowel contexts.

	/s/		/ʃ/	
	Front-vowel context	**Back-vowel context**	**Front-vowel context**	**Back-vowel context**
CI	6.7 kHz (1.46 kHz)	6.65 kHz (1.59 kHz)	5.31 kHz (1.24 kHz)	5.12 kHz (1.21 kHz)
NH	8.04 kHz (1.5 kHz)	7.89 kHz (1.5 kHz)	5.45 kHz (1.14 kHz)	5.17 kHz (1.12 kHz)

Model 3a (reference: CI, /ʃ/, back-vowel context) showed a significant main effect of Target Consonant ($\hat{\beta}_2$ = 1.33, *SE* = 0.24, *t* = 5.54, $p < 0.001$) and a significant interaction between Group and Target Consonant ($\hat{\beta}_4$ = 1.29, *SE* = 0.31, *z* = 4.08, $p < 0.01$). The main effect of Target Consonant indicates that in back-vowel contexts, children with CIs produced /s/ with higher centroid frequency than /ʃ/. The interaction indicates that the change in centroid frequency across target consonants in back-vowel contexts was larger for children with NH than children with CIs.

Model 3b (reference: CI, /s/, front-vowel context) showed significant main effects of Group ($\hat{\beta}_1$ = 1.43, *SE* = 0.25, *t* = 5.70, $p < 0.001$) and Target Consonant ($\hat{\beta}_2$ = -1.26, *SE* = 0.23, *t* = -5.39 $p < 0.01$), as well as a significant interaction between Group and Target Consonant ($\hat{\beta}_4$ = -1.26, *SE* = 0.31, *t* = -4.02, $p < 0.001$). The main

effect of Group indicates that in front-vowel contexts, children with NH produced /s/ with higher centroid frequency than children with CIs. The main effect of Target Consonant indicates that children with CIs produced /ʃ/ with lower centroid frequency than /s/ in front-vowel contexts. The interaction indicates that the change in centroid frequency across consonants in front-vowel contexts was larger for children with NH than for children with CIs.

Model 3c (reference: NH, /ʃ/, front-vowel context) showed a significant main effect of Target Consonant ($\hat{\beta}_2$ = 2.53, *SE* = 0.21, *t* = 12.10 *p* < 0.001). The significant interaction between Group and Target Consonant is identical to that of Model B3. The main effect of Target Consonant indicates that children with NH produced /s/ with significantly higher centroid frequency than /ʃ/ in front-vowel contexts. The effect of Vowel Context showed a non-significant but noteworthy trend ($\hat{\beta}_3$ = -0.31, *SE* = 0.13, *t* = -2.44, *p* = 0.015), suggesting that children with NH produced /ʃ/ with slightly lower centroid frequency in back-vowel contexts compared to front-vowel contexts.

Model 3d (reference: NH, /s/, back-vowel context) showed significant main effects of Group ($\hat{\beta}_1$ = -1.32, *SE* =0.25, *t* = -5.37, *p* < 0.001) and Target Consonant ($\hat{\beta}_2$ = -2.62, *SE* = 0.21, *t* = -12.68, *p* < 0.01). The significant interaction between Group and Target Consonant is identical to that of model A3. The effect of Group indicates that children with CIs produced /s/ in back-vowel contexts with lower centroid frequency than children with NH. The effect of Target Consonant indicates

that in back-vowel contexts, children with NH produced /ʃ/ with lower centroid frequency than /s/.

Taken together, these results show that across vowel contexts, all children produced /s/ with higher centroid frequency than /ʃ/. However, the magnitude of this change was not equal across groups: children with CIs produced /s/ and /ʃ/ with less acoustic differentiation than children with NH. While both groups produced equivalent centroid frequencies for /ʃ/, children with CIs consistently produced /s/ with lower centroid frequencies than children with NH.

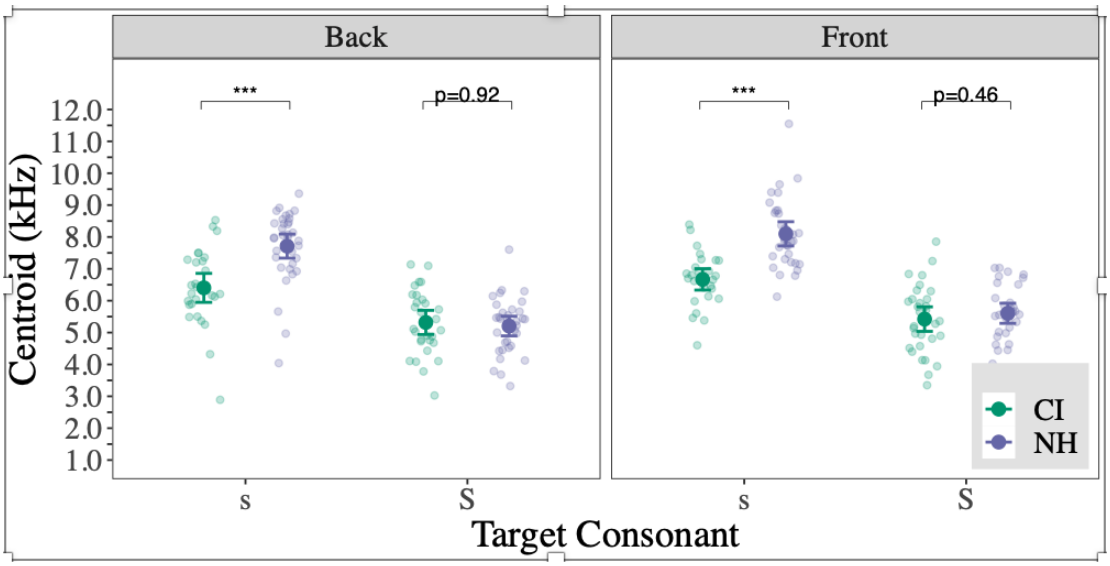


Figure 3. Mean centroid frequencies for each group (large circles) with ±2 standard error bars and individual data (smaller transparent circles) for word-initial /s/ and /ʃ/ tokens in both back- and front-vowel contexts. *** indicates $p < 0.001$

Robustness of Contrast Results: Adults

The model that classified tokens as /s/ or /ʃ/ was extremely accurate: out of 913 tokens, only four were classified incorrectly (1 /ʃ/ and 3 /s/ tokens), yielding an overall robustness of contrast of 99.6% for the adults. Results from model 4a indicate

that classification accuracy was significantly higher for the /s/-/ʃ/ contrast compared to the /t/-/k/ contrast ($\hat{\beta}_1$ = -2.49, *SE* = 0.52, *z* = -4.76, $p < 0.01$), where accuracy overall was 95.0% for the same subset of adults. This trend can be seen in Figures 4 and 5. These figures show centroid frequency on the x-axis and speakers on the y-axis, and each dot represents a production that is color-coded based on the target consonant. These figures are also faceted by vowel context. Each speaker has 2 rows, because the adults completed two sessions with different Wordlists. The x-axis is labeled by each individual speaker's session ID and the percentage of their tokens from that session that were correctly classified by the model. These figures clearly illustrate the increased discriminability for /s/ and /ʃ/ tokens compared to /t/ and /k/ tokens, and that the difference across contrasts is driven by a systematic increase in the centroid frequency for /k/ in front vowel contexts.

Power Analysis

Data from the adults were used to conduct a simulation-based power analysis using the model described in Chapter 2, Equation (1) for classifying tokens and the *mixedPower* package in R (Kumle, Võ, & Draschkow, 2021). The purpose of this analysis was to determine the probability of detecting a large enough effect of centroid to accurately classify consonants across a range of sample sizes. The first step in estimating power was to simulate 1000 new data sets for several different sample sizes. Three sample sizes were selected based on the number of sessions that were available for robustness of contrast analyses: n = 32 sessions for adults, n = 44 sessions for the children's /s/-/S/ analysis, and n = 64 for the children's /t/-/k/ analysis. After the data were simulated, the model was refit to the new data to

generate parameter estimates, and each effect was tested for significance. Power was then calculated as the number of simulations where the effect was significant divided by the total number of simulations.

The output of the model that classified tokens as /s/ or /S/ showed no significant main effect nor significant interaction. However, even though these effects were not statistically significant, they were sufficiently large to be relevant for the current purpose: the percentage of tokens correctly classified by this model as /s/ or /S/ was >99%. Thus, even though the lowest *z*-value from the model output was 1.17, the critical *z*-value for testing significance in the simulations was set to a more conservative 1.65. The results of this simulation-based power analysis revealed that a sample size of 32 yielded 99% power to detect a sufficiently-sized main effect of centroid, 65% power for a main effect of vowel context, and 62% power for the interaction. A sample size of 44 yielded 99% power to detect a main effect of centroid, 76% power for a main effect of vowel context, and 75% power for the interaction. A sample size of 64 yielded 100% power to detect a main effect of centroid, 86% power for a main effect of vowel context, and 85% power for the interaction. A simulation-based power analysis was also conducted for the adults' /t/-/k/ data using the same sample sizes and the same critical *z*-value. Sample sizes of 32, 44, and 64 all showed 100% power for detecting the main effects of centroid and vowel context, and 25%, 34%, and 45% power for detecting a significant interaction.

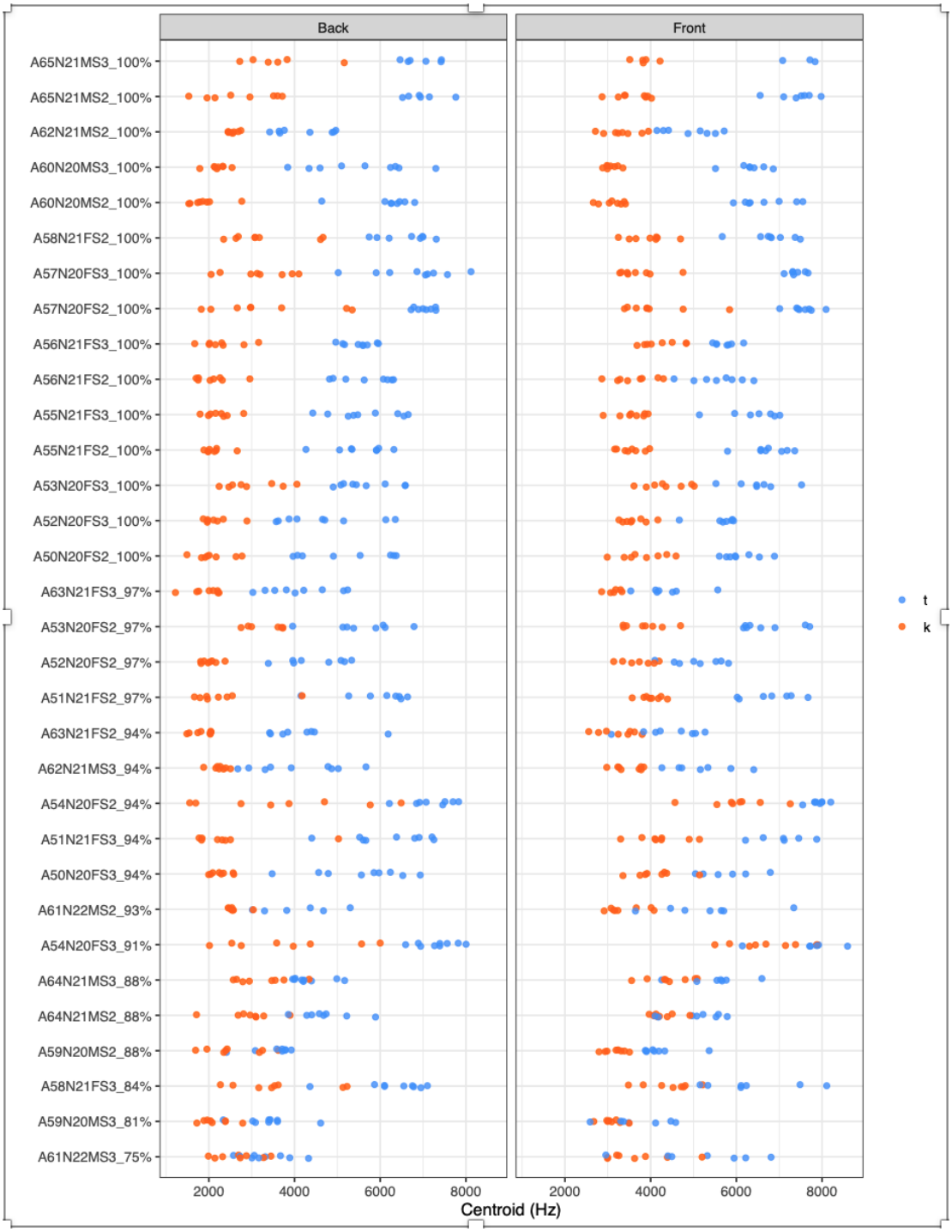


Figure 4. Adults' robustness of /t/-/k/ contrast. Centroid frequency for each adults' /t/ (blue dots) and /k/ (orange dots) productions in both front- and back-vowel contexts, arranged in descending order of robustness of contrast, quantified by the percentage of tokens correctly classified by the model. Each adult participated twice, and sessions are presented separately along the x-axis. Each adult's unique 3-digit ID number includes their age (in years), sex (M or F), and the Wordlist used to elicit tokens.

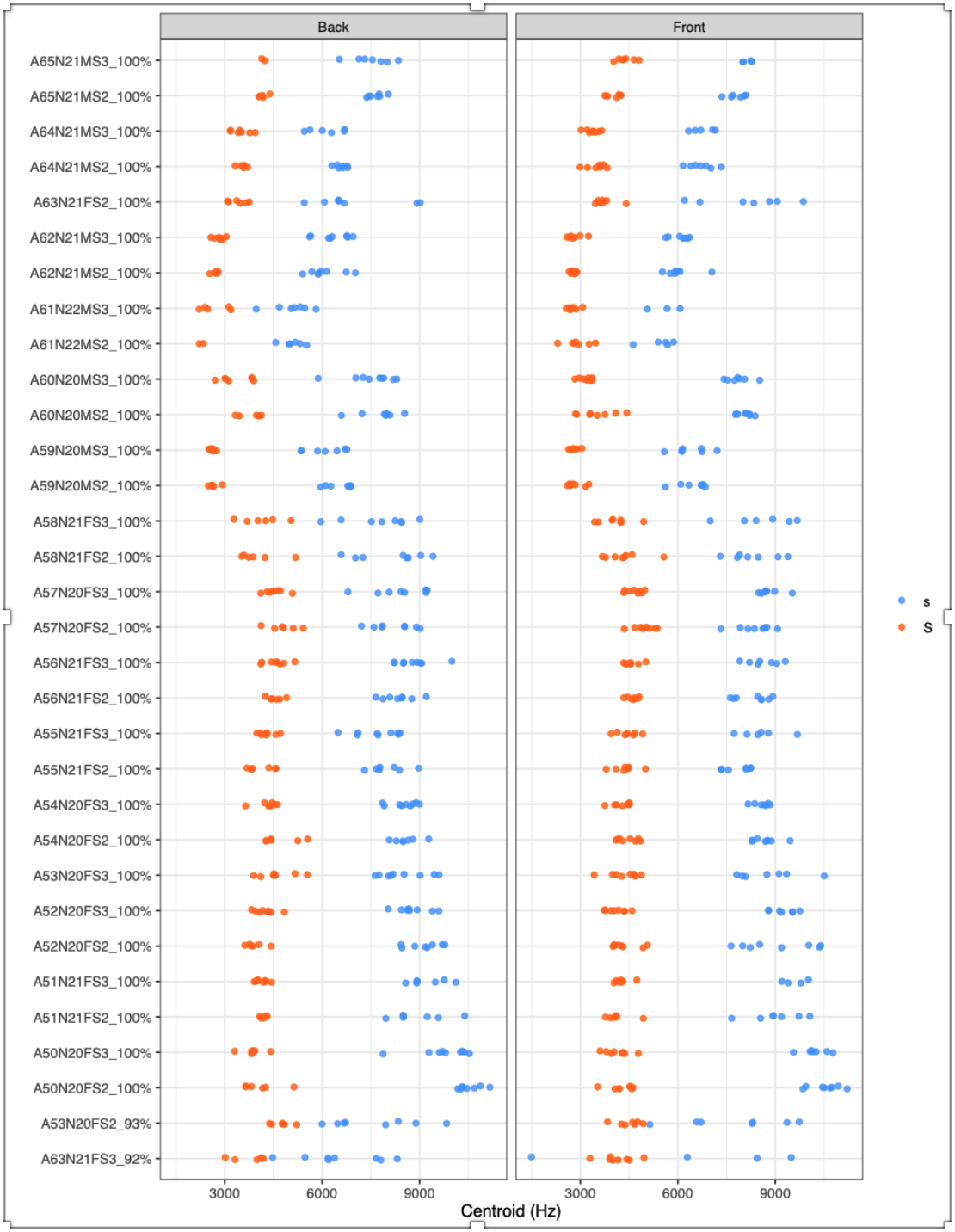


Figure 5. Adults' robustness of /s/-/ʃ/ contrast. Centroid frequency for each adults' /s/ (blue dots) and /ʃ/ (orange dots) productions in both front- and back-vowel contexts, arranged in descending order of robustness of contrast, quantified by the percentage of tokens correctly classified by the model. Each adult participated twice, and sessions are presented separately along the x-axis. Each adult's unique 3-digit ID number includes their age (in years), sex (M or F), and the Wordlist used to elicit tokens.

Robustness of Contrast Results: Adults vs. Children

For the /t/-/k/ contrast, results of model 4b predicting classification accuracy based on Group showed that adults produced a more robust contrast than children with NH ($\hat{\beta}_1$ = -1.71, *SE* = 0.23, *z* = -7.58, *p* < 0.001) and children with CIs ($\hat{\beta}_2$ = -1.65, *SE* = 0.23, *z* = -7.21, *p* < 0.001). For the /s/-/ʃ/ contrast, results from model 4c showed a similar trend, with significantly higher robustness for adults compared to children with NH ($\hat{\beta}_1$ = -3.34, *SE* = 0.63, *z* = -5.32, *p* < 0.001) and children with CIs ($\hat{\beta}_2$ = -4.39, *SE* = 0.63, *z* = -7.02, *p* < 0.001). Accuracy of predictions is shown for adults and children across all target consonants and vowel contexts in Table 5.

Table 5. Overall robustness of the /t/-/k/ and /s/-/ʃ/ contrasts for adults and children, and the accuracy of predictions by target consonant and vowel context.

Group	**Target /t/**		**Target /k/**		**Target /s/**		**Target /ʃ/**	
	Front	**Back**	**Front**	**Back**	**Front**	**Back**	**Front**	**Back**
Adults	94% (24%)	95% (21%)	95% (22%)	94% (24%)	99% (10%)	100% (6%)	100% (0%)	100% (7%)
Children with NH	53% (50%)	91% (29%)	81% (39%)	85% (36%)	84% (37%)	92% (27%)	94% (25%)	87% (34%)
Children with CI	48% (50%)	87% (34%)	82% (39%)	96% (20%)	77% (42%)	83% (438%)	82% (39%)	71% (46%)

Robustness of Contrast Results: Children with CIs vs. Children with NH

Results of 4d (reference: CI, /t/-/k/) showed neither a significant main effect of Group or Contrast, indicating that children with CIs and children with NH

produced equally robust /t/-/k/ contrasts, and that children with CIs produced equally robust /t/-/k/ and /s/-/ʃ/ contrasts. The interaction between Group and Contrast was significant ($\hat{\beta}_3$ = 1.08, *SE* = 0.31, *z* = 3.5, *p* < 0.001), indicating that the (lack of) difference between groups for robustness of the /t/-/k/ contrast was not similar for the /s/-/ʃ/ contrast. The results of model 4e (reference: NH, /s/-/ʃ/) showed significant main effects of Group ($\hat{\beta}_1$ = -0.99, *SE* = 0.29, *z* = -3.37, *p* < 0.001), Contrast ($\hat{\beta}_2$ = -1.00, *SE* = 0.23, *z* = -4.24, *p* < 0.001), and the significant interaction is identical to that of the first Model. The main effect of Group indicates that children with CIs produced less robust /s/-/ʃ/ contrasts compared to children with NH. The main effect of Contrast indicates that Children with NH produced less robust /t/-/k/ contrasts compared to /s/-/ʃ/ contrasts.

Taken together, these results show that children with CIs produced equally robust contrasts. Children with NH produced more robust /s/-/ʃ/ contrasts compared to /t/-/k/ contrasts, and their /s/-/ʃ/ contrasts were more robust than the /s/-/ʃ/ contrasts produced by children with CIs. These trends are illustrated in Figures 6 and 7.

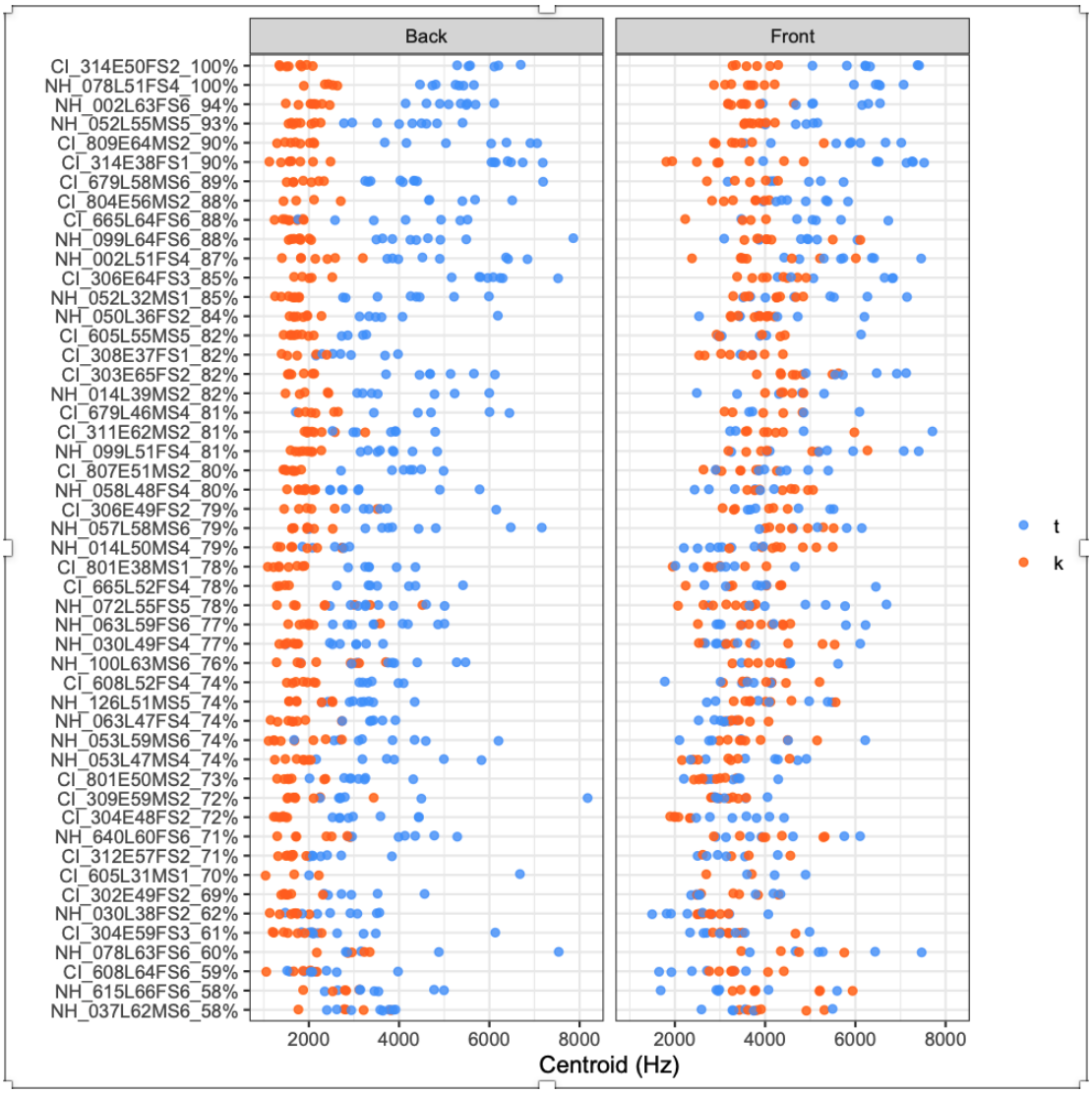


Figure 6. Children's robustness of /t/-/k/ contrast. Centroid frequency for each child's /t/ (blue dots) and /k/ (orange dots) productions in both front- and back-vowel contexts, arranged in descending order of robustness of contrast, quantified by the percentage of tokens correctly classified by the model. Each child's unique 3-digit ID number is prepended by their hearing status and includes their age at test (in months), and sex (M or F).

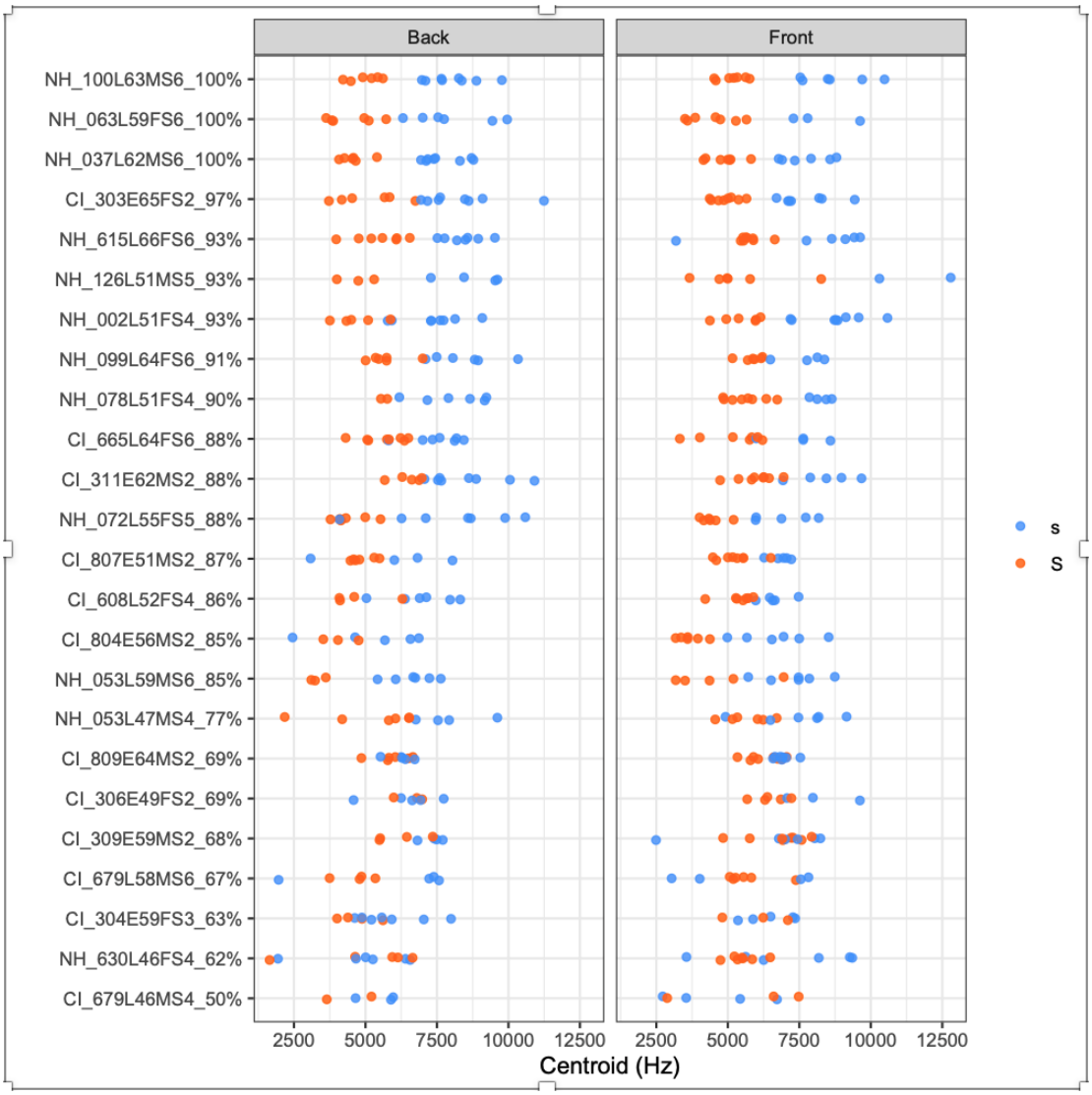


Figure 7. Children's robustness of /s/-/ʃ/ contrast. Centroid frequency for each child's /s/ (blue dots) and /ʃ/ (orange dots) productions in both front- and back-vowel contexts, arranged in descending order of robustness of contrast, quantified by the percentage of tokens correctly classified by the model. Each child's unique 3-digit ID number is prepended by their hearing status and includes their age at test (in months), and sex (M or F).

Results from Chapter 3 showed that children with CIs produced /k/ in front-vowel context with lower centroid frequency compared to children with NH. One possible explanation for this finding is that children with CIs do not produce the palatal allophone of /k/ in front-vowel contexts, but children with NH do. Figures 8

and 9 provide some evidence of this possibility. Figure 8 shows the distribution of adults' /k/ and /t/ centroids for front-vowel context (purple) and back-vowel context (green), with shading to indicate frequencies that were common across vowel contexts. The bi-modal distribution for /k/ (but not /t/) illustrates that adults produce two distinct allophones of /k/, and the velar /k/s produced in back-vowel contexts have lower centroid frequencies than the palatal /k/s in front-vowel contexts. Figure 9 similarly shows the distribution of /k/ and /t/ centroids for children with and without CIs. Both groups of children also show bi-modal distributions for /k/, indicating they produce the same /k/ allophones as adults. However, children with CIs produced /k/ in front-vowel contexts with lower centroid frequency than their peers, thereby reducing the degree of separation between their /k/ allophones.

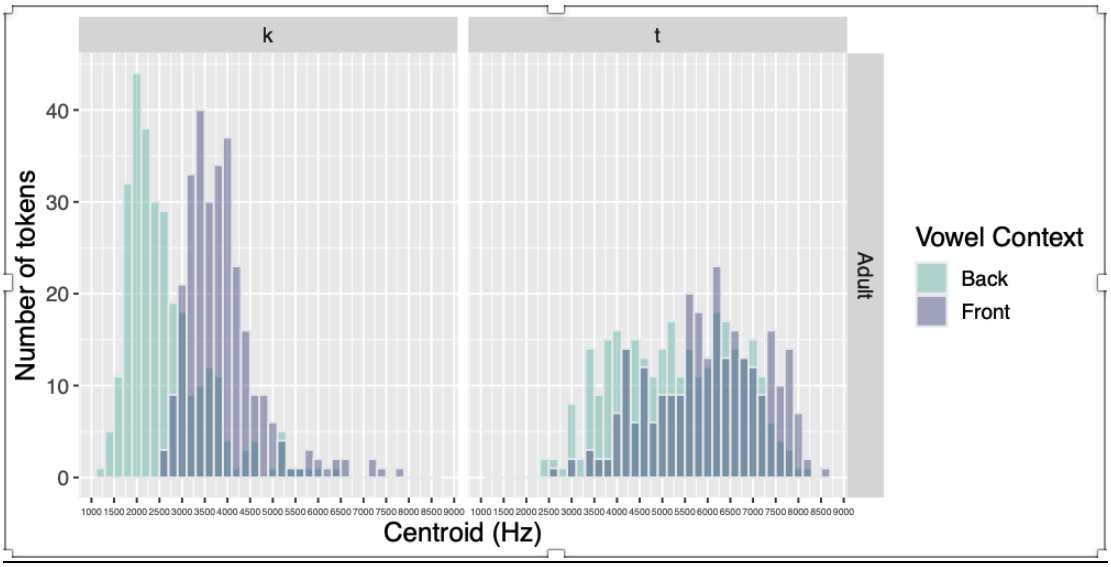


Figure 8. Distribution of centroid frequencies for adults' productions of /k/ and /t/ across vowel contexts.

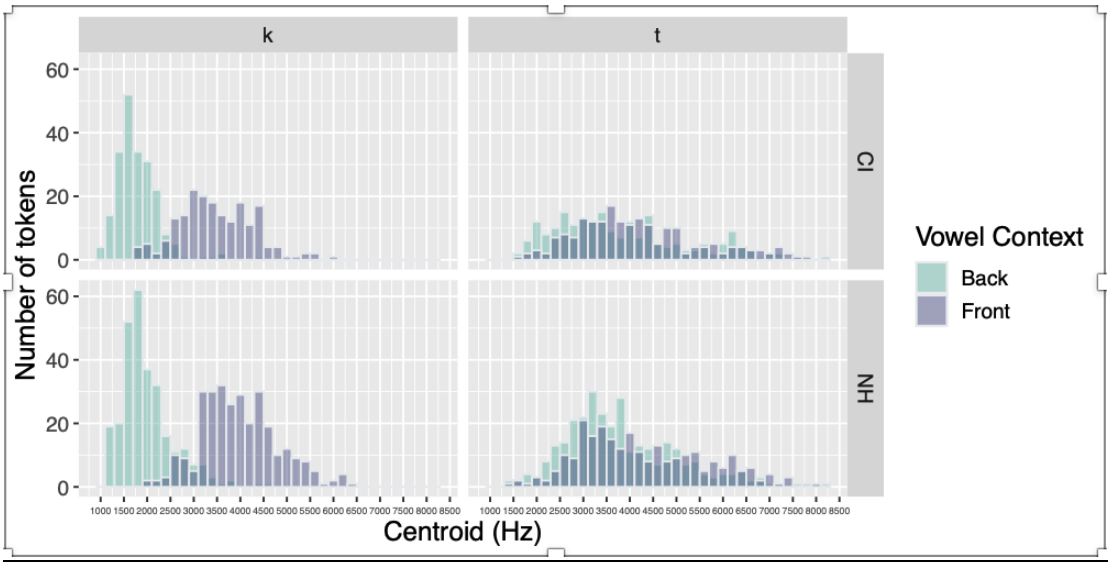


Figure 9. Distribution of centroid frequencies for children with CIs' and children with NH's productions of /k/ and /t/ across vowel contexts.

Discussion

The primary aim of this paper was to compare accuracy, error patterns, and spectral features of word-initial /s/ and /ʃ/ produced by 3- to 5-year-old children with and without CIs. A secondary aim was to compare robustness of both the /t/-/k/ and the /s/-/ʃ/ contrasts across the two groups of children and adults. These analyses were intended to elucidate the effects of signal degradation on speech acquisition for a late-acquired fricative contrast that is not only cued spectrally, but also requires more advanced speech-motor control.

As predicted, children with CIs produced /s/ less accurately than their peers with NH in both front- and back-vowel contexts. Contrary to predictions, though, /ʃ/ was produced with equal accuracy across groups and contexts. Furthermore, an interaction showed that in front-vowel contexts, the difference in accuracy across consonants was not equal across groups: while children with NH produced /s/ more

accurately than /ʃ/, children with CIs produced /s/ less accurately than /ʃ/. This trend was present in back-vowel contexts as well, but the effect was not significant when using the adjusted alpha-level of $p = 0.0125$. Contrary to the findings for /t/ and /k/ (Chapter 3, Johnson et al., 2021), there were no changes in accuracy within or across groups due to vowel context. One explanation for this finding is that the longer duration of fricatives provides enough steady-state information to glean cues to consonant identity regardless of vowel context. Any acoustic consequences of coarticulation likely occur near the end of the fricative, and the effect on perception at that point is negligible.

The difference in accuracy of /s/ productions across groups—particularly in the absence of differences for /ʃ/ accuracy—provides strong evidence that limitations of CI speech processing programs impact speech acquisition. If children with CIs demonstrated a simple delay due to auditory deprivation, they would be more likely to produce both consonants less accurately than their peers. Furthermore, differences in accuracy across consonants were not significant, suggesting that the level of neuro-motor control required to produce /s/ and /ʃ/ is similar. Thus, the difference in accuracy across groups for /s/ is more likely to be related to auditory-perceptual factors rather than speech-motor constraints. Information in the noise spectrum used to identify /s/ is not only distorted, but also frequencies above 8.0 kHz are eliminated. To put this number in perspective, adults in this study produced /s/ with a mean centroid frequency of 7.9 kHz (sd = 1.43 kHz) and children with NH produced /s/ with a mean centroid frequency of 8.0 kHz (sd = 1.50 kHz). Thus, children with CIs

do not have access to the high-frequency spectral information necessary to accurately perceive /s/.

Children with CIs and children with NH were equally likely to produce manner and place errors, and none of the children produced voicing errors. Both groups of children produced /s/ with higher centroid frequency than /ʃ/. However, in both front- and back-vowel contexts, children with NH produced /s/ with higher centroid frequency than children with CIs, and the difference between /s/ and /ʃ/ centroids in front-vowel contexts was larger for children with NH than children with CIs. Similar to the results for accuracy, the fact that group differences in centroid frequency were observed for /s/ productions—but not /ʃ/ productions—support the claim that degraded auditory input, particularly for sounds with high-frequency spectral cues that extend beyond the upper frequency limit of CIs, impacts acquisition of spectral contrasts for children using CIs. These results provide additional evidence that a degraded auditory input impacts acquisition of spectral contrasts for children with CIs.

In this chapter, I found that adults produced a more robust /s/-/ʃ/ contrast than /t/-/k/ contrast, i.e., centroid frequency measured at fricative midpoint more reliably differentiated /s/ from /ʃ/ across speakers and vowel contexts compared to centroid frequency of burst spectra calculated for /t/ and /k/. This finding is likely driven by relatively different effects of coarticulation for stops and fricatives. When produced in front-vowel contexts, the centroid frequency for the release burst of /k/ is impacted by coarticulation to a greater extent than the centroid frequency at fricative midpoint /ʃ/. The systematic increase in centroid frequency for /k/ in front-vowel contexts reduces

spectral distance and increases overlap between /k/ and /t/ categories, thus making them less discriminable. Fricatives are much longer in duration than stops. Although the spectra for /s/ and /ʃ/ also vary due to coarticulation, these effects are more likely to appear near the offset of the fricative and have little impact on the noise spectrum at the midpoint.

Although using a single static measure is ideal in terms of simplicity and clinical feasibility, it is possible that a different measure (e.g., an additional static measure, or dynamic measures) are needed to better differentiate /t/ and /k/ across vowel contexts, even within-speaker. In this study, the standard deviation of centroid frequencies for adults' /t/ productions was 1.32 kHz for /t/, but only 1.07 kHz for /k/ productions, despite the systematic variation of /k/ across vowel contexts. Because centroid is a measure of overall energy concentration, it may be less valid to measure sounds like /t/ that have relatively diffuse spectra. Alternatively, the 4th spectral moment, kurtosis, is a measure of peakedness, which may be more useful for differentiating /t/ and /k/ across vowel contexts.

Even though adults produced less a robust /t/-/k/ than /s/-/ʃ/ contrast, both contrasts were more robust than the contrasts produced by children. This finding replicates previous work reported by Nittouer (1995). The current study also confirmed the utility of calculating robustness of contrast compared to using a simple measure of accuracy: even though children produced sounds as "correct" when judged by transcription analyses, measurable acoustic differences were observed in their productions. Using a more fine-grained, continuous measure to characterize the speech of children allows researchers and clinicians to better understand how speech

is fine-tuned during the stages of development between consonant emergence, consonant proficiency, and adult-like mastery.

The current study also extended previous work on contrast acquisition to younger children. An early study by Nittrouer (1995) found that adults produced a greater difference between centroid frequencies for /s/ and /ʃ/ compared to children, and there were no differences between groups of children who were 3, 5, and 7 years old. Similarly, Nissen and Fox (2005) found that children ages 3-4 did not produce a contrast between /s/ and /ʃ/, but a contrast emerged for 5-year-old children, even though it was still less robust than adults' contrast. This study confirmed that 3- to 5-year-old children do indeed produce a spectral contrast between /s/ and /ʃ/, even though it is not as robust as adults'. One reason that younger children in the current study produced a contrast may be related to differing demands of the speech elicitation procedures. Previous studies elicited consonant tokens in carrier phrases: "This is a __," and "It's a __ Bob," and they elicited tokens using only picture prompts (Nissen & Fox, 2015; Nittrouer, 1995). By contrast, the protocol used in the current study was designed to reduce both speech and language demands by eliciting words in isolation and providing not only a picture prompt, but also an auditory model of each word. Repeating words given an auditory model and reducing the added articulatory demands of embedding target consonants in carrier phrases likely reduced variability within and across speakers in our study, thereby making effects across groups more detectable. (Unfortunately, previous studies did not report standard deviations, so it is not possible to compare variability across studies directly.)

Comparing across groups of children with and without CIs, the current study extended previous work on the /s/-/ʃ/ to younger children and also to a new contrast. The results in this study were similar to previous studies: children with CIs produced less robust /s/-/ʃ/ contrasts compared to their peers with NH, and the reduced spectral distance was driven by lower centroid frequencies for /s/, even though there were no differences in centroid for /ʃ/. However, robustness of the /t/-/k/ contrast was similar across groups. Although children with CIs produced /k/ with systematically lower frequency in front-vowel contexts compared to their peers with NH, this did not result in more robust /t/-/k/ contrasts. One explanation for this finding is that centroid frequency for /t/ was highly variable. This may be due to the fact that the spectrum for /t/ is relatively diffuse, and therefore not well-characterized by centroid measurements.

Although centroid frequency is a fairly useful continuous measure for quantifying robustness of these two place contrasts, there may be better alternatives that can also be used for contrasts that are not easily differentiated using a single acoustic measure. For example, rating tokens on a visual analog scale (VAS) would also provide a continuous measure upon which robustness of contrast could be calculated. This approach may also be more ecologically valid, because not only is the ultimate goal understanding children's speech, but listeners are able to use all of the cues available in the rich acoustic signal to judge tokens, whereas obtaining a single acoustic measurement like centroid is reductionist. Furthermore, centroid frequency is not invariant, and therefore it may not translate to perception in a very meaningful way. Transcribing tokens as intermediate, as was done in this study,

provides more information, but the output of this procedure is still more categorical than if tokens were judged on a continuous VAS. It is possible that covert contrasts could be quantified more readily using VAS than either transcription or acoustic measures (e.g., Munson et al., 2011). Additionally, previous work has shown that acoustic robustness of contrast is related to perceptual judgments of consonants for adult speakers (Newman et al., 2001), children with NH and typical development (Holliday et al., 2015), and children using CIs (Reidy et al., 2017). If the goal of quantifying robustness of contrast is to examine its effects on listener's perception of the speech, it may be more useful to derive a measure of robustness from a perceptual measure rather than an acoustic one.

Although previous work by Holliday et al. (2015) showed that a more robust acoustic contrast between /s/ and /ʃ/ tokens produced by children was associated with higher goodness ratings by listeners, it is possible that this relationship is more complex. As noted above, children with CIs coarticulated less and produced /k/ with lower centroid frequency in front-vowel contexts compared to their peers, which increased the spectral distance between /t/ and /k/, yielding a more robust contrast. It is possible that in this case, increased robustness of contrast would not correspond to perceived goodness in the same way that a more robust /s/-/ʃ/ contrast did in Holliday et al. (2015). It may be that listeners *expect* to hear a more fronted /k/ in front-vowel context, and they would in fact perceive these less-fronted /k/ tokens produced by children with CIs in front-vowel contexts as poor exemplars. Perhaps this measure of contrast is not directly related to perceptual goodness because these productions violate expectations for listeners. Future work is needed to assess the perceptual

consequences of the acoustic differences found for /k/ in this study to determine if they negatively impact children's speech intelligibility.

There are also many allophones for /t/ in English. It is possible that word-initial aspirated /t/ is uniquely challenging for children with CIs to learn. The aspiration that occurs following the release burst has a noise spectrum similar to that of fricatives. Thus, children must be sensitive not only to spectral place cues, but also temporal and amplitude cues in order to differentiate aspirated /t/ from the affricate "ch." Although temporal envelope information is well-preserved when processed through a CI, it is possible that the sampling rate of 4ms is insufficient to convey extremely short, but potentially critical, information in the signal, such as the amplitude-rise time necessary for children using CIs to differentiate affricates from aspirated stops.

In conclusion, this study showed how acquisition of spectral contrasts for young children learning English is impacted by the speech processing limitations of CIs. In particular, differences across groups in terms of both accuracy and centroid frequency for /s/, but not /ʃ/, suggest that the high-frequency spectral information used to perceive /s/ and differentiate it from /ʃ/ is not readily available for children with CIs. Speech pathologists must consider the unique auditory-perceptual limitations of CIs when assessing speech, choosing speech targets, and designing interventions for children who use CIs. For example, it may be useful for speech pathologists to target /ʃ/ or other, more visible fricatives before /s/ so that children can make use of their auditory-articulatory/proprioceptive feedback loop while their

speech-motor system is still maturing, and then progress to targeting /s/ once children have demonstrated the ability to produce fricative manners of articulation.

Chapter 5: General Discussion

The purpose of this dissertation was to examine the role of signal degradation in phonological acquisition for children with CIs. Without a doubt, the period of auditory deprivation prior to cochlear implantation (usually at least a year) contributes to delays in speech development for this group of children. In this study, two place-of-articulation contrasts produced by children with CIs were compared to those produced by age- and sex-matched peers with NH. Overall, children with CIs produced all sounds except /ʃ/ less accurately than their peers with NH. Children with NH who have the same amount of hearing experience (often referred to as "hearing-age matches") were not used as a comparison group, because they would be younger than the children with CIs, and therefore would have less mature speech-motor skills and smaller oral cavities.

A number of findings from this dissertation suggest that signal degradation influences phonological acquisition for children with CIs. The two contrasts investigated in this study are both differentiated by spectral cues. These contrasts were chosen precisely because they were likely to be highly impacted by signal degradation and result in atypical patterns of acquisition for children with CIs compared to children with NH, and this prediction was borne out.

The sounds that diverged most from the typical sequence of development for children with CIs were sounds with the highest-frequency cues. For children with CIs, the sound produced with the lowest accuracy is the sound with the highest frequency information, /s/. Children with CIs produced /ʃ/ with similar accuracy levels as their peers, but they produced /s/ less accurately than /ʃ/, while their peers produced /s/

more accurately than /ʃ/. Acoustically, children with CIs also had less robust /s/-/ʃ/ contrasts compared to the children with NH.

Similarly, for the /t/ vs. /k/ contrast, although both groups of children produced /t/ with equal accuracy to /k/, children with NH produced both of these sounds near ceiling-levels of accuracy (> 90%), indicating they had already acquired this contrast, while children with CIs were still in the process of acquiring these sounds (67% accuracy). Therefore, it is more informative to compare the children with CIs in this study to children with NH in normative samples. Smit et al. (1990) report that word-initial /t/ is acquired earlier than word-initial /k/ for children with typical speech development. When considering this developmental trend, children with CIs demonstrated a relative difference in their acquisition of /t/ and /k/: at this stage of development, /t/ should have higher accuracy than /k/. In this case, it is unlikely that accuracy of /k/ is relatively boosted, especially when considering the impact of vowel context. Thus, the relatively lower accuracy for /t/ for children with CIs may be attributed to higher spectral energy in the /t/ burst. Contrary to prediction, the difference in centroids between /t/ and /k/ in front-vowel contexts was significantly greater for children with CIs compared to children with NH. This increase in spectral distance between /t/ and /k/ was the result of children with CIs producing a less fronted /k/ in front-vowel contexts compared to their peers. Although this lack of coarticulation in front-vowel contexts resulted in a more robust contrast, the consequences of these acoustic differences are unknown. Future research is needed to determine if listeners expect a higher-frequency release burst for /k/ before front vowels, and whether the speech of children who do not produce this variation is

perceived as less intelligible. Additional work is also necessary to determine if younger children with NH also coarticulate less during an earlier stage of development, prior to having fully acquired this contrast.

There are a number of limitations to this study. First, the research assistants who transcribed the productions were not blind to the child's hearing status or age, which may be a source of bias. Even though reliability among transcribers was acceptable, it is possible everyone imposed similar biases while transcribing. It would be beneficial for these productions to be re-transcribed by naïve listeners who know nothing about the children. Not only would this process increase the reliability of transcriptions, but it would also provide valuable information about tokens that are inherently difficult to transcribe, particularly the intermediate tokens, distortions, and erroneous productions.

Second, the current study used an objective measure of voicing for /t/ and /k/, but it would be useful to determine whether these tokens were perceived by listeners as their voiced counterparts. Although voice-onset-time is a relatively reliable cue to voicing, there are other cues that lead to the perception of a voiced or voiceless stop. It is possible that the objective measure characterized some productions as voiced even though they would be perceived as voiceless and vice versa.

Third, vowel context was coded as front or back based on the target vowels. Because the vowels were not transcribed, it is possible that children produced the wrong vowels. For example, perhaps children with CIs produced /k/ with lower centroid frequency in front-vowel contexts because they were substituting back vowels.

Finally, as noted above, this study used a comparison group of peers with NH who were matched in terms of chronological age rather than a group of children with the same amount of hearing experience. Although a hearing-age group would better control for the period of auditory deprivation, there are several critical challenges to this approach for the current study. Primarily, children with the same hearing age would be at least one year younger than the children with CIs. Not only would these younger children have less mature motor skills, they would also have smaller oral cavities and shorter vocal tracts, so acoustic features could not be directly compared across groups. Furthermore, the experimental task is too demanding for children under age three, and several of the hearing-age matches would be two or younger. If it were feasible in future research to shorten the task or collect data across several testing sessions, it would be worthwhile to obtain similar data in younger children and compare children with CIs to hearing-age matches. This type of comparison could be useful to determine, for example, whether younger children with NH progress through a similar stage while acquiring /k/ where they do not produce the allophonic variation of /k/ in front-vowel contexts. Studies on even younger children could elucidate the developmental progression of acquiring not only the /t/-/k/ contrast, but also the progression towards adult-like speech in terms of producing allophones.

This dissertation answered several questions related to speech development in children with CIs, but there is room for additional work that could help elucidate the process of learning. Phonetic categories are abstract. There is no single acoustic exemplar of a phonetic category, even within a language, because there is so much variability within and across speakers. Thus, while the current work employed an

outcome-driven approach to describe learning of phonetic categories, a mechanism-driven approach, such as the one recently described by Schatz et al. (2021) may be especially useful when considering the effects of highly degraded auditory speech input for children. We presume that children are learning phonetic categories or phonemic contrasts during speech development, and we test these assumptions by quantifying the extent to which children's perceptions or productions approximate adult's categories. However, there are limitations to imposing adults' knowledge when investigating what children are learning, and there are far-reaching implications if infants go through an initial learning process that does not involve learning phonetic categories (Feldman, Goldwater, Dupoux, & Schatz, 2021). Schatz et al. (2021) describe an alternative, mechanism-driven approach to study how children may learn phonetic categories. Using a model of learning that is given realistic speech input, built on large-scale simulations across two languages (English and Japanese), the authors found that infants may learn phonetic categories via distributional (i.e., statistical) learning, but the units gleaned from the distributions of natural speech are shorter and acoustically too fine-grained to correspond to adults' phonetic categories. It would be interesting to apply this modeling framework to children using CIs: when given realistic input (i.e., speech that is systematically and highly degraded in a way that reflects the output of CI speech processing algorithms), what are the units that children are able to learn?

Another direction for future research on the speech development in children with CIs is to study affricates in greater depth. Affricates are more complex than either stops or fricatives, and they are differentiated from other manners of

articulation by a complex combination of cues associated with temporal, amplitude, *and* spectral features. Stopping is a common phonological process that occurs when children are learning to produce fricatives. The children with NH in this study produced affricate substitutions for fricatives, reflecting a process of stopping while transitioning to the correct manner of articulation. Stopping is a process common during fricative acquisition that makes sense from a motor development perspective, because stops are easier to produce than fricatives and can be used as a springboard to articulating fricatives produced at the same place. However, the children with CIs in this study produced affricate substitutions for both target stops and target fricatives. Producing manner-of-articulation errors where the output is more complex than the target was not expected. It is possible that the temporal information transmitted by CIs is not as robust at necessary to differentiate among all manners of articulation. The distinction between fricatives, stops, and affricates require attention to very short, fine-grained temporal cues in combination with both spectral and amplitude cues. Although producing target affricates was not a part of this study, comparing the affricate substitutions produced by children with CIs to target affricate productions would be useful in determining whether they are producing covert contrasts. If children can perceive these more fine-grained temporal cues that differentiate affricates from stops and fricatives, they are more likely to produce covert contrasts. Alternatively, if children are not able to perceive these cues, they are more likely to produce no contrast between target and substitute affricates. Another approach to answering this question could be to conduct studies investigating how children with and without CIs weight acoustic cues to affricate-manner distinctions (e.g., Giezen,

Escudero, & Baker, 2010; Nittrouer & Lowenstein, 2015). More salient cues are likely to be weighted stronger (Nittrouer, Tarr, Moberly, & Lowenstein, 2014). Cue-weighting studies that incorporate tests of auditory sensitivity in addition to perceptual attention to acoustic cues would be particularly useful to determine whether or not children have access to the fine-grained temporal, spectral, and amplitude information necessary to distinguish affricates, and if so, can attention to these cues be improved through auditory training (e.g., Moberly, Lowenstein, & Nittrouer, 2016).

Regardless of the limitations in scope and methodology, there are clear implications of this research relevant to several disciplines involved in maximizing outcomes for people using CIs, including otologic surgeons, engineers, audiologists, and speech-language pathologists. Most CIs are programmed using a universal approach, where frequency bands are assigned to electrodes using fixed presets that are not customized to the specific listener. However, a more individualized approach to programming based may be warranted. A promising new strategy is *image-guided CI programming* (IGCIP) technique, which uses post-operative CT scans to determine the precise locations of intracochlear electrodes (and the existence of extracochlear cochlear electrodes) to customize programming (Noble, Labadie, Gifford, & Dawant, 2013). Through IGCIP, frequency bands are reallocated, and specific electrodes are de/activated based on objective measurements of the relative position of each electrode and its contact with the auditory nerve.

In terms of improving CI processing strategies, a marked improvement in speech recognition was observed with the invention of the multi-channel CI, which

capitalized on the tonotopic organization of the cochlea and central auditory structures to transmit place cues. Speech processing algorithms that transmit information relevant to place features, especially at higher frequencies, have been shown specifically to improve consonant identification (Skinner et al., 1999). It is possible that children require information at even higher frequencies than adults in order to learn speech sounds like /s/. Children in this study produced /s/ with centroid frequencies ranging as high as 12,800 Hz. Without information above 8,000 Hz, children may be unable to effectively use auditory feedback from their own productions to adjust their articulatory gestures and refine their speech towards progressively more intelligible, adult-like output.

In terms of facilitating phonological development, speech-language pathologists would be remiss to treat children with CIs simply as younger children with NH: there are fundamental differences in auditory experience that create unique constraints on learning phonemes by listening, mimicking, and adjusting based on auditory-perception and articulatory feedback loops. Children with CIs are likely to need explicit training to master place-of-articulation contrasts that are signaled by spectral cues, especially when sounds are not visually salient. Providing children with multiple cues (proprioceptive, phonetic placement, visual) should be helpful when the degraded signal is less useful for auditory feedback. Incorporating cues from a variety of sensory systems may be particularly useful for children who had the least amount of auditory input prior to implantation. Compared to children with a progressive loss, children born with profound, bilateral hearing loss who received minimal benefit from hearing aids prior to implantation are more likely to have undergone cortical

changes, such as cross-modal reorganization and neural pruning in the auditory processing centers, which occurs after prolonged absence of meaningful input (e.g., Lee et al., 2001; Moore & Linthicum, 2007). Speech-language pathologists should also not expect children with CIs to follow the same developmental path in terms of order of acquisition of phonemes as children with NH when planning which sounds and contrasts to target in therapy. In addition to a different order-of-acquisition of sounds, the error patterns produced by children with CIs while acquiring phonemes are different from those produced by children with NH, and this difference should be taken into account when determining prognosis for improvement and planning treatment. Finally, the reduced robustness of the /s/ vs. /ʃ/ contrast within *correct* productions for children with CIs suggest that speech-language pathologists may need to continue working on contrasts even after individual sounds are perceived as correct to improve intelligibility. Adopting clinically feasible continuous measures, such as rating productions on a visual-analog scale, may be useful for quantifying robustness of contrast and monitoring progress.

To conclude, the method used in this study – to focus on multiple productions of a small number of sounds – is incredibly useful for examining the developmental process of phonological acquisition for children with CIs. As we wait for future advances in the cochlear implant devices and programming strategies, the goal of speech pathologists is to figure out the optimal strategies for teaching children to listen and produce intelligible speech given the devices that are available today.

Appendix A: Detailed Model Outputs from Chapter 2

Table A1. Results from Model 1, a logistic mixed-effects model predicting the log-likelihood that an adult's production was accurately classified as /t/ or /k/ with the reference category: Centroid (ERB_N).

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	3.44	0.25	13.72	< 0.001*
Representation	-0.02	0.19	-0.10	0.92
Measure	-0.75	0.17	-4.52	< 0.001*
Representation x Measure	-0.08	0.23	-0.35	0.72

Table A2. Results from Model 2, a logistic mixed-effects model predicting the log-likelihood that an adult's production was accurately classified as /t/ or /k/ with the reference category: Peak (Hz).

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	2.58	0.23	11.13	< 0.001*
Representation	0.10	0.14	0.72	0.47
Measure	0.83	0.16	5.09	< 0.001*
Representation x Measure	-0.08	0.23	-0.35	0.72

Appendix B: Detailed Audiologic History for Children with CIs

ID	Sex	Etiology	Age at Onset (months)	Age at Activation (months)	Age at Test (months)	Device	Bi/Uni/ Bimodal
300E	M	Genetic	0	13	57	Med EL Opus 2	Bilateral; simultaneous
301E	F	Unknown	0	45; 49	53	Med EL Opus 2	Bilateral; R/L
302E	F	Unknown	0	13; 16	37; 49	AB Neptune (later Naida)	Bilateral; R/L
303E	F	Unknown	6	13	65	Med EL Opus 2	Bilateral; simultaneous
304E	F	Genetic	0	12; 13	48; 59	Med EL Opus 2	Bilateral; R/L
305E	F	Unknown	0	28; 39	44; 56	AB Neptune	Bilateral; R/L
306E	F	Unknown	0	11; 38	49; 64	Med EL Opus 2	Bilateral; R/L
307E	M	Genetic	0	15; 16	44	Cochlear Nucleus 5	Bilateral; R/L
308E	F	Genetic	0	13	37	Med EL Opus 2	Bilateral; simultaneous
309E	M	Genetic	0	7; 7	59	Cochlear Nucleus 6	Bilateral; simultaneous
310E	F	Genetic	0	23	51	Cochlear Nucleus 6	Bilateral; simultaneous
311E	M	Unknown	9	13; 53	62	Advanced Bionics Harmony	Bilateral; L/R
312E	F	Genetic	0	48	44; 57	AB Neptune	Unilateral; R Bilateral (R/L)

314E	F	Unknown	10	17; 24	38; 50	AB Neptune (later Naida)	Bilateral; R/L
605L	M	Unknown	0	Unknown	31; 43; 55	Med EL Opus 2	BiModal CI/L, HA/R
608L	F	Genetic	0	9	39; 52; 64	Cochlear Nucleus 5	Bilateral; simultaneous
665L	F	Genetic	0	12; 17	52; 64	Med EL Opus 2	Bilateral; R/L
679L	M	Genetic	0	29	34; 46; 58	Cochlear Nucleus 6	BiModal; CI/L; HA/R
800E	M	Genetic	30	37; 38	65	Med EL Opus 2	Bilateral; simultaneous
801E	M	Unknown	1	15; 20	38; 50	AB Neptune	Bilateral; simultaneous (w/ complication requiring 2nd implantation of R)
803E	F	Unknown	0	33	41	Cochlear Nucleus 6	BiModal; CI/L; HA/R
804E	M	Genetic	0	7	56	Cochlear Nucleus 5	Bilateral; simultaneous
806E	M	Genetic	14	34	42	Cochlear Nucleus 6	Unilateral (L)
807E	M	Genetic (progressive)	6	22	51	Cochlear Nucleus 5	BiModal; CI/R; HA/L
808E	F	Genetic	0	6	37	Cochlear Nucleus 5	Bilateral; simultaneous
809E	M	Meningitis	6	8; 32	64	Cochlear Nucleus 5	Bilateral; R/L

Appendix C: Detailed Model Outputs from Chapter 3

Table C1. Results of Model 1a predicting accuracy based on Target Consonant, Group, and Vowel Context, with reference category: CI, /k/, back-vowel context.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	1.22	0.35	3.45	< 0.001*
Group	2.11	0.54	3.87	< 0.001*
Target Consonant	-0.10	0.31	-0.34	0.73
Vowel Context	-0.68	0.23	-3.00	0.003*
Group x Target Consonant	0.44	0.53	0.82	0.41
Group x Vowel Context	0.95	0.42	2.28	0.02
Target Consonant x Vowel Context	0.71	0.32	2.23	0.03
Group x Target Consonant x Vowel Context	-1.55	0.59	-2.62	0.009*

Table C2. Results of Model 1b predicting accuracy based on Target Consonant, Group, and Vowel Context, with reference category: CI, /t/, front-vowel context.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	1.15	0.32	3.57	< 0.001*
Group	1.94	0.50	3.90	< 0.001*
Target Consonant	-0.61	0.30	-2.06	0.04
Vowel Context	-0.03	0.23	-0.14	0.89
Group x Target Consonant	1.11	0.52	2.12	0.03
Group x Vowel Context	0.60	0.42	1.43	0.15
Target Consonant x Vowel Context	0.71	0.32	2.23	0.03
Group x Target Consonant x Vowel Context	-1.55	0.59	-2.62	0.009*

Table C3. Results of Model 1c predicting accuracy based on Target Consonant, Group, and Vowel Context, with reference category: NH, /k/, front-vowel context.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	3.59	0.43	8.32	< 0.001*
Group	-3.05	0.55	-5.54	< 0.001*
Target Consonant	-0.50	0.44	-1.13	0.26
Vowel Context	-0.26	0.35	-0.76	0.45
Group x Target Consonant	1.11	0.52	2.12	0.03
Group x Vowel Context	0.95	0.42	2.28	0.02
Target Consonant x Vowel Context	0.83	0.50	1.68	0.09
Group x Target Consonant x Vowel Context	-1.55	0.59	-2.62	0.009*

Table C4. Results of Model 1d predicting accuracy based on Target Consonant, Group, and Vowel Context, with reference category: NH, /t/, back-vowel context.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	3.66	0.41	8.86	< 0.001*
Group	-2.55	0.52	-4.92	< 0.001*
Target Consonant	-0.34	0.45	-0.74	0.46
Vowel Context	-0.57	0.35	-1.61	0.11
Group x Target Consonant	0.44	0.53	0.82	0.41
Group x Vowel Context	0.60	0.42	1.43	0.15
Target Consonant x Vowel Context	0.83	0.50	1.68	0.09
Group x Target Consonant x Vowel Context	-1.55	0.59	-2.62	0.009*

Table C5. Results of Model 2a predicting manner errors based on Group and Target Consonant, with reference category: CI, target /t/.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	-0.64	0.24	-2.64	0.008*
Group	-3.43	1.09	-3.16	0.002*
Target Consonant	-0.73	0.24	-3.05	0.002*
Group x Target Consonant	2.23	1.20	1.86	0.06

Table C6. Results of Model 2b predicting manner errors based on group and target consonant. Reference category: NH, target /k/.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	-2.57	0.59	-4.32	< 0.001*
Group	1.19	0.64	1.88	0.06
Target Consonant	-1.50	1.17	-1.28	0.20
Group x Target Consonant	2.23	1.20	1.86	0.06

Table C7. Results of Model 2c predicting voicing errors based on group and target consonant. Reference category: CI, target /t/.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	-1.87	0.46	-4.05	< 0.001*
Group	2.86	0.79	3.63	< 0.001*
Target Consonant	-0.13	0.29	-0.45	0.66
Group x Target Consonant	-1.79	0.79	-2.26	0.024*

Table C8. Results of Model 2d predicting voicing errors based on group and target consonant. Reference category: NH, target /k/.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	-0.93	0.62	-1.48	0.14
Group	-1.07	0.74	-1.44	0.15
Target Consonant	1.92	0.73	2.62	0.009*
Group x Target Consonant	-1.79	0.79	-2.26	0.024*

Table C9. Results of Model 2e predicting place errors based on group and target consonant. Reference category: CI, target /t/.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	-0.38	0.23	-1.68	0.09
Group	-0.59	0.48	-1.24	0.21
Target Consonant	1.09	0.23	4.80	< 0.001*
Group x Target Consonant	-0.08	0.56	-0.13	0.89

Table C10. Results of Model 2f predicting place errors based on group and target consonant. Reference category: NH, target /k/.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	0.04	0.38	0.11	0.91
Group	0.67	0.44	1.52	0.13
Target Consonant	-1.01	0.51	-2.00	0.05
Group x Target Consonant	-0.08	0.56	-0.13	0.89

Table C11. Results of Model 3a predicting centroid frequency based on Target Consonant, Group, and Vowel Context, with reference category: CI, /k/, back-vowel context.

	Estimate	**Std. Error**	***t*-value**	***p*-value**
Intercept	1.79	0.07	25.81	< 0.001*
Group	0.21	0.09	2.26	0.025
Target Consonant	1.97	0.15	12.77	< 0.001*
Vowel Context	1.69	0.09	19.61	< 0.001*
Group x Target Consonant	-0.15	0.21	-0.72	0.47
Group x Vowel Context	0.29	0.11	2.62	0.009*
Target Consonant x Vowel Context	-1.36	0.12	-11.16	< 0.001*
Group x Target Consonant x Vowel Context	-0.16	0.16	-0.99	0.32

Table C12. Results of Model 3b predicting centroid frequency based on Target Consonant, Group, and Vowel Context, with reference category: CI, /t/, front-vowel context.

	Estimate	**Std. Error**	***t*-value**	***p*-value**
Intercept	4.09	0.16	26.35	< 0.001*
Group	0.19	0.21	0.91	0.37
Target Consonant	-0.61	0.16	-3.93	< 0.001*
Vowel Context	-0.33	0.09	-3.83	< 0.001*
Group x Target Consonant	0.31	0.21	1.45	0.15
Group x Vowel Context	-0.13	0.11	-1.20	0.23
Target Consonant x Vowel Context	-1.36	0.12	-11.16	< 0.001*
Group x Target Consonant x Vowel Context	-0.16	0.16	-0.99	0.32

Table C13. Results of Model 3c predicting centroid frequency based on Target Consonant, Group, and Vowel Context, with reference category: NH, /k/, front-vowel context.

	Estimate	**Std. Error**	***t*-value**	***p*-value**
Intercept	3.98	0.06	66.75	< 0.001*
Group	-0.50	0.09	-5.35	< 0.001*
Target Consonant	0.31	0.14	2.15	0.03
Vowel Context	-1.98	0.07	-27.91	< 0.001*
Group x Target Consonant	0.31	0.21	1.45	0.15
Group x Vowel Context	0.29	0.11	2.62	0.009*
Target Consonant x Vowel Context	1.51	0.10	14.88	< 0.001*
Group x Target Consonant x Vowel Context	-0.16	0.16	-0.99	0.32

Table C14. Results of Model 3d predicting centroid frequency based on Target Consonant, Group, and Vowel Context, with reference category: NH, /t/, back-vowel context.

	Estimate	**Std. Error**	***t*-value**	***p*-value**
Intercept	3.82	0.14	26.55	< 0.001*
Group	-0.06	0.21	-0.28	0.78
Target Consonant	-1.82	0.14	-12.89	< 0.001*
Vowel Context	0.46	0.07	6.36	< 0.001*
Group x Target Consonant	-0.15	0.21	-0.72	0.47
Group x Vowel Context	-0.13	0.11	-1.20	0.23
Target Consonant x Vowel Context	1.51	0.10	14.88	< 0.001*
Group x Target Consonant x Vowel Context	-0.16	0.16	-0.99	0.32

Appendix D: Detailed Model Outputs from Chapter 4

Table D1. Results of Model 1a predicting accuracy based on Target Consonant, Group, and Vowel Context, with reference category: CI, /ʃ/, back-vowel context.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	0.53	0.41	1.28	0.20
Group	0.48	0.57	0.83	0.41
Target Consonant	-0.88	0.36	-2.43	0.02
Vowel Context	0.02	0.23	0.10	0.92
Group x Target Consonant	1.22	0.50	2.41	0.02
Group x Vowel Context	-0.29	0.32	-0.92	0.36
Target Consonant x Vowel Context	-0.07	0.31	-0.21	0.83
Group x Target Consonant x Vowel Context	0.40	0.44	0.91	0.36

Table D2. Results of Model 1b predicting accuracy based on Target Consonant, Group, and Vowel Context, with reference category: CI, /s/, front-vowel context.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	-0.39	0.33	-1.20	0.23
Group	1.80	0.47	3.85	< 0.001*
Target Consonant	0.94	0.36	2.64	< 0.001*
Vowel Context	0.04	0.21	0.20	0.84
Group x Target Consonant	-1.62	0.50	-3.22	< 0.001*
Group x Vowel Context	-0.11	0.30	-0.35	0.72
Target Consonant x Vowel Context	-0.07	0.31	-0.21	0.83
Group x Target Consonant x Vowel Context	0.40	0.44	0.91	0.36

Table D3. Results of Model 1c predicting accuracy based on Target Consonant, Group, and Vowel Context, with reference category: NH, /ʃ/, front-vowel context.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	0.73	0.40	1.85	0.06
Group	-0.18	0.57	-0.32	0.75
Target Consonant	0.67	0.35	1.92	0.05
Vowel Context	0.27	0.22	1.23	0.22
Group x Target Consonant	-1.62	0.50	-3.22	0.001*
Group x Vowel Context	-0.29	0.32	-0.92	0.36
Target Consonant x Vowel Context	-0.33	0.31	-1.09	0.28
Group x Target Consonant x Vowel Context	0.40	0.44	0.91	0.36

Table D4. Results of Model 1d predicting accuracy based on Target Consonant, Group, and Vowel Context, with reference category: NH, /s/, back-vowel context.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	1.34	0.32	4.15	<0.001*
Group	-1.69	0.46	-3.67	<0.001*
Target Consonant	-0.34	0.35	-0.97	0.33
Vowel Context	0.06	0.21	0.30	0.77
Group x Target Consonant	1.22	0.50	2.41	0.016
Group x Vowel Context	-0.11	0.30	-0.35	0.72
Target Consonant x Vowel Context	-0.33	0.31	-1.09	0.28
Group x Target Consonant x Vowel Context	0.40	0.44	0.91	0.36

Table D5. Results of Model 2a predicting manner errors based on Group and Target Consonant, with reference category: CI, target /s/.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	0.75	0.38	2.00	0.05
Group	0.11	0.56	0.20	0.84
Target	0.10	0.25	0.41	0.68
Group x Target	-0.33	0.39	-0.85	0.40

Table D6. Results of Model 2b predicting manner errors based on Group and Target Consonant, with reference category: NH, target /ʃ/.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	0.64	0.40	1.58	0.11
Group	0.22	0.57	0.38	0.70
Target	0.23	0.30	0.76	0.45
Group x Target	-0.33	0.39	-0.85	0.40

Table D7. Results of Model 3a predicting centroid frequency based on Target Consonant, Group, and Vowel Context, with reference category: CI, /ʃ/, back-vowel context.

	Estimate	**Std. Error**	***t*-value**	***p*-value**
Intercept	5.18	0.18	29.55	< 0.001*
Group	0.03	0.24	0.11	0.92
Target Consonant	1.33	0.24	5.54	< 0.001*
Vowel Context	0.17	0.14	1.15	0.25
Group x Target Consonant	1.29	0.32	4.08	< 0.001*
Group x Vowel Context	0.14	0.19	0.75	0.45
Target Consonant x Vowel Context	-0.06	0.21	-0.30	0.76
Group x Target Consonant x Vowel Context	-0.03	0.27	-0.10	0.92

Table D8. Results of Model 3b predicting centroid frequency based on Target Consonant, Group, and Vowel Context, with reference category: CI, /s/, front-vowel context.

	Estimate	**Std. Error**	***t*-value**	***p*-value**
Intercept	6.61	0.19	34.57	< 0.001*
Group	1.43	0.25	5.70	< 0.001*
Target Consonant	-1.26	0.23	-5.39	< 0.001*
Vowel Context	-0.10	0.15	-0.68	0.50
Group x Target Consonant	-1.26	0.31	-4.02	< 0.001*
Group x Vowel Context	-0.12	0.20	-0.60	0.55
Target Consonant x Vowel Context	-0.06	0.21	-0.30	0.76
Group x Target Consonant x Vowel Context	-0.03	0.27	-0.10	0.92

Table D9. Results of Model 3c predicting centroid frequency based on Target Consonant, Group, and Vowel Context, with reference category: NH, /ʃ/, front-vowel context.

	Estimate	**Std. Error**	***t*-value**	***p*-value**
Intercept	5.52	0.16	35.36	< 0.001*
Group	-0.17	0.23	-0.75	0.46
Target Consonant	2.53	0.21	12.10	< 0.001*
Vowel Context	-0.31	0.13	-2.44	0.015
Group x Target Consonant	-1.26	0.31	-4.02	< 0.001*
Group x Vowel Context	0.14	0.19	0.75	0.45
Target Consonant x Vowel Context	0.09	0.18	0.52	0.61
Group x Target Consonant x Vowel Context	-0.03	0.27	-0.10	0.92

Table D10. Results of Model 3d predicting centroid frequency based on Target Consonant, Group, and Vowel Context, with reference category: NH, /s/, back-vowel context.

	Estimate	**Std. Error**	***t*-value**	***p*-value**
Intercept	7.83	0.16	49.45	< 0.001*
Group	-1.32	0.25	-5.37	< 0.001*
Target Consonant	-2.62	0.21	-12.68	< 0.001*
Vowel Context	0.22	0.12	1.79	0.07
Group x Target Consonant	1.29	0.32	4.08	< 0.001*
Group x Vowel Context	-0.12	0.20	-0.60	0.55
Target Consonant x Vowel Context	0.09	0.18	0.52	0.61
Group x Target Consonant x Vowel Context	-0.03	0.27	-0.10	0.92

Table D11. Results of Model 4a predicting classification accuracy of adults' production based on Contrast, with reference level /s/-/ʃ/.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	5.98	0.59	10.17	< 0.001*
Contrast	-2.49	0.52	-4.76	< 0.001*

Table D12. Results of Model 4b predicting classification accuracy of /t/ and /k/ tokens produced by adults, children with NH, and children with CIs based on Group, with adults as the reference category.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	3.05	0.18	17.36	<0.001*
Group (CI)	-1.65	0.23	-7.21	<0.001*
Group (NH)	-1.71	0.23	-7.58	<0.001*

Table D13. Results of Model 4c predicting classification accuracy of /s/ and /ʃ/ tokens produced by adults, children with NH, and children with CIs based on Group, with adults as the reference category.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	5.76	0.56	10.22	<0.001*
Group (CI)	-4.39	0.63	-7.02	<0.001*
Group (NH)	-3.34	0.63	-5.32	<0.001*

Table D14. Results of Model 4d predicting classification accuracy of children's production based on Group, Contrast, and the interaction between Group and Contrast, with reference level CI, /t/-/k/.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	1.37	0.13	10.75	<0.001
Group	-0.09	0.18	-0.51	0.61
Contrast	-0.09	0.19	-0.45	0.66
Group x Contrast	1.08	0.31	3.52	<0.001

Table D15. Results of Model 4e predicting classification accuracy of children's production based on Group, Contrast, and the interaction between Group and Contrast, with reference level NH, /s/-/ʃ/.

	Estimate	**Std. Error**	***z*-value**	***p*-value**
Intercept	2.28	0.23	9.96	<0.001
Group	-0.99	0.29	-3.37	<0.001
Contrast	-0.99	0.23	-4.24	<0.001
Group x Contrast	1.08	0.31	3.52	<0.001

Appendix E: List of Target Words

Appendix E. List of target words (written orthographically and phonemically) and their vowel contexts, and the number of times each word was elicited within Wordlists 1-3.

Word	IPA transcription	Vowel Context	Wordlist: repetitions
Cake	/keɪk̟/	Front	1: 2 2: 1 3: 1
Candle	/kændl̩/	Front	1: 0 2: 1 3: 1
Candy	/kændi/	Front	1: 2 2: 1 3: 2
Car	/kɑɹ/	Back	1: 2 2: 1 3: 0
Cat	/kæt/	Front	1: 2 2: 1 3: 1
Catch	/kætʃ̟/	Front	1: 0 2: 1 3: 0
Coat	/koʊ̯t/	Back	1: 2 2: 1 3: 1
Coffee	/kɑfi/	Back	1: 0 2: 1 3: 1
Comb	/koʊ̯m/	Back	1: 0 2: 1 3: 1
Cookie	/kʊki/	Back	1: 2 2: 2 3: 2
Cousin	/kʌzɪn/	Back	1: 0 2: 0 3: 1
Cup	/kʌp/	Back	1: 2 2: 1 3: 1
Cutting	/kʌɾɪŋ/	Back	1: 0 2: 1

			3: 1
Keys	/kiz/	Front	1: 0 2: 1 3: 1
Kitchen	/kɪtʃɪn/	Front	1: 2 2: 1 3: 1
Kitten	/kɪtn̩/	Front	1: 0 2: 0 3: 1
Kitty	/kɪɾi/	Front	1: 2 2: 1 3: 0
Table	/teɪbl̩/	Front	1: 2 2: 1 3: 1
Take	/teɪk/	Front	1: 0 2: 1 3: 1
Tape	/teɪp/	Front	1: 2 2: 1 3: 0
Teacher	/titʃɹ/	Front	1: 0 2: 0 3: 1
Teddy bear	/tɛdibeɪɹ/	Front	1: 2 2: 2 3: 2
Tent	/tɛnt/	Front	1: 0 2: 2 3: 2
Tickle	/tɪkl̩/	Front	1: 2 2: 1 3: 1
Tiger	/taɪgɻ/	Back	1: 0 2: 1 3: 1
Toast	/toʊst/	Back	1: 2 2: 1 3: 1
Toaster	/toʊstɻ/	Back	1: 0 2: 1 3: 2
Tongue	/tʌŋ/	Back	1: 2 2: 2

			3: 1
Tooth	/tuθ/	Back	1: 2 2: 1 3: 0
Toothbrush	/tuθbɹʌʃ/	Back	1: 0 2: 1 3: 1
Towel	/taʊl̩/	Back	1: 0 2: 0 3: 1
Tummy	/tʌmi/	Back	1: 2 2: 1 3: 2
Share	/ʃe͡ɪɹ/	Front	1: 4 2:1 3: 0
Sharing	/ʃe͡ɪɹɪŋ/	Front	1: 0 2: 2 3: 1
Sheep	/ʃip/	Front	1: 4 2: 2 3: 2
Shell	/ʃɛl/	Front	1: 0 2: 1 3: 1
Ship	/ʃɪp/	Front	1:0 2: 2 3: 2
Shovel	/ʃʌvl/	Back	1: 2 2:1 3:2
Shoe	/ʃu/	Back	1: 4 2: 1 3: 0
Shoes	/ʃuz/	Back	1:0 2: 1 3: 1
Shower	/ʃa͡ʊwɚ/	Back	1: 2 2:1 3:2
Sugar	/ʃʊ͡gɚ/	Back	1: 0 2: 0 3: 1
Shadow	/ʃædoʊ͡/	Front	1: 0

			2: 0 3: 2
Cereal	/siɹiəl/	Front	1: 0 2: 0 3: 2
Sad	/sæd/	Front	1: 2 2: 2 3: 0
Sandwich	/sændwɪt͡ʃ/	Front	1: 2 2: 1 3: 1
Sandbox	/sændbɑks/	Front	1: 0 2: 1 3: 1
Sick	/sɪk/	Front	1: 2 2: 1 3: 0
Sidewalk	/sa͡ɪdwɑk/	Front	1: 0 2: 1 3: 1
Sink	/sɪŋk/	Front	1: 0 2: 1 3: 1
Sister	/sɪstə˞/	Front	1:0 2: 1 3: 1
Scissors	/sɪzə˞z/	Front	1: 2 2: 1 3: 1
Seven	/sɛvɪn/	Front	1: 0 2: 0 3: 1
Sock	/sɑk/	Back	1: 2 2: 1 3: 0
Soup	/sup/	Back	1: 2 2: 1 3: 2
Soap	/so͡ʊp/	Back	1: 2 2: 1 3: 0
Suitcase	/sutke͡ɪs/	Back	1: 0 2: 2 3: 2
Sun	/sʌn/	Back	1: 2

			2: 1 3: 2
Sunny	/sʌni/	Back	1: 0 2: 1 3: 0
Summer	/sʌmɚ/	Back	1: 0 2: 0 3: 1

Bibliography

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *lme4: linear mixed-effects models using Eigen and S4*. R package version 1.1-11.

Bates D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. doi:10.18637/jss.v067.i01

Bass-Ringdahl, S. (2010). The relationship of audibility and the development of canonical babbling in young children with hearing impairment. *The Journal of Deaf Studies and Deaf Education, 15*(3), 287-310.

Baudonck, N., Van, L., D'haeseleer, E., & Dhooge, I. (2011). A comparison of the perceptual evaluation of speech production between bilaterally implanted children, unilaterally implanted children, children using hearing aids, and normal-hearing children. *International Journal of Audiology, 50*(12), 912-919. doi:10.3109/14992027.2011.605803

Behrens, S. J., & Blumstein, S. E. (1988). Acoustic characteristics of English voiceless fricatives: a descriptive analysis. *Journal of Phonetics, 16*(3), 295–298. https://doi.org/10.1016/S0095-4470(19)30504-2

Blamey, P. J., Barry, J., & Jacq, P. (2001). Phonetic inventory development in young cochlear implant users 6 years postoperation. *Journal of Speech, Language, and Hearing Research : Jslhr, 44*, 73–79.

Blumstein, S. E. & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of the Acoustical Society of America*, *66*(4), 1001-1017.

Brauer, M., & Curtin, J. (2017). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*.

doi:10.1037/met0000159

Boersma, P. & Weenink, D. (2018). *Praat: doing phonetics by computer*. [Computer program]. Version 6.0.30. http://www.praat.org/

Crowe, K., & McLeod, S. (2014). A systematic review of cross-linguistic and multilingual speech and language outcomes for children with hearing loss. *International Journal of Bilingual Education and Bilingualism*, *17*(3), 287–309. https://doi.org/10.1080/13670050.2012.758686

Crowe, K., & McLeod, S. (2020). Children's English consonant acquisition in the United States: A review. *American Journal of Speech-Language Pathology*, *29*(4), 2155–2169. https://doi.org/10.1044/2020_AJSLP-19-00168

Cychosz, M., Munson, B., & Edwards, J. (2021). Practice and experience predict coarticulation in child speech. *Language Learning and Development.* https://doi.org/10.1080/15475441.2021.1890080

Donaldson, G., & Kreft, H. (2006). Effects of vowel context on the recognition of initial and medial consonants by cochlear implant users. *Ear & Hearing, 27*(6), 658-677. doi:10.1097/01.aud.0000240543.31567.54

Dunn, L.M., & Dunn, D.M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition*. San Antonio, TX: Pearson Education.

Edwards, J., Gibbon, F., & Fourakis, M. (1997). On discrete changes in the acquisition of the alveolar/velar stop consonant contrast. *Language and Speech, 40*(2), 203-210.

Edwards, J., & Beckman, M. (2008). Methodological questions in studying consonant acquisition. *Clinical Linguistics & Phonetics, 22*(12), 937-956.

Eilers, R., & Oller, D. (1994). Infant vocalizations and the early diagnosis of severe hearing impairment. *Journal of Pediatrics, 124*(2), 199-203.

Ertmer, D. J., & Goffman, L. (2011). Speech production accuracy and variability in young

cochlear implant recipients: comparisons with typically developing age-peers. *Journal of Speech, Language, and Hearing Research: JSLHR*, *54*(1), 177–89. https://doi.org/10.1044/1092-4388(2010/09-0165)

Ertmer, D. J., Kloiber, D. T., Jung, J., Kirleis, K. C., & Bradford, D. (2012). Consonant production accuracy in young cochlear implant recipients: developmental sound classes and word position effects. *American Journal of Speech-Language Pathology*, *21*(4), 342–53. https://doi.org/10.1044/1058-0360(2012/11-0118)

Faes, J., & Gillis, S. (2016). Word initial fricative production in children with cochlear implants and their normally hearing peers matched on lexicon size. *Clinical Linguistics & Phonetics*, *30*(12), 959–982. https://doi.org/10.1080/02699206.2016.1213882

Faes, J., Gillis, J., & Gillis, S. (2016). Phonemic accuracy development in children with cochlear implants up to five years of age by using levenshtein distance. *Journal of Communication Disorders*, *59*, 40–58. https://doi.org/10.1016/j.jcomdis.2015.09.004

Feldman, N. H., Goldwater, S., Dupoux, E., & Schatz, T. (2021). Do infants really learn phonetic categories? *Open Mind*, *5*, 113-131. https://doi.org/10.1162/opmi_a_00046

Fenson L, Marchman VA, Thal DJ, Dale PS, Reznick JS, Bates E. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual.* 2. Baltimore, MD: Brookes.

Fitzpatrick, E., Olds, J., Gaboury, I., McCrae, R., Schramm, D., & Durieux-Smith, A. (2012). Comparison of outcomes in children with hearing aids and cochlear implants. *Cochlear Implants Inernational, 13*(1), 5-15. doi:10.1179/146701011X12950038111611

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America*, *84*(1), 115-123.

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, Y. B. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America*, *110*(2), 1150–1163.

Giezen, M. R., Escudero, P., & Baker, A. (2010). Use of acoustic cues by children with cochlear implants. *Journal of Speech, Language, and Hearing Research: JSLHR*, *53*(6), 1440–1457.

Gillis, S., Schauwers, K., & Govaerts, P. (2002). *Babbling milestones and beyond*. In K. Schauwers, P. Govaerts, & S. Gillis (Eds.), Language acquisition in young children with a cochlear implant (pp. 23–40). Antwerp, Belgium: University of Antwerp.

Glasberg, B. & Moore, B. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1), 103-138. doi:10.1016/0378-5955(90)90170-T

Goldman, R., & Fristoe, M. (2000*). Goldman–Fristoe Test of Articulation—2*. Circle Pines, MN: American Guidance Services.

Heinz, J. M., & Stevens, K. N. (1961). On the properties of voiceless fricative consonants. *The Journal of the Acoustical Society of America, 33*(5), 589–596. https://doi.org/10.1121/1.1908734

Holliday, J., Reidy, P., Beckman, M., & Edwards, J. (2015). Quantifying the robustness of the English sibilant fricative contrast in children. *Journal of Speech, Language, and Hearing Research: JSLHR*, *58*(3), 622-637. doi:10.1044/2015_JSLHR-S-14-0090

Hughes, G. W., & Halle, M. (1956). Spectral properties of fricative consonants. *The Journal of the Acoustical Society of America*, *28*(2), 303–310. https://doi.org/10.1121/1.1908271

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434-446.

Johnson, A.A., Reidy, P.F., & Edwards, J. (2018). Quantifying robustness of the /t/-/k/ contrast using a single, static spectral feature. *Journal of the Acoustical Society of America, 144*(2), E105-E111.

Johnson, A.A., Bentley, D.M, Munson, B., & Edwards, J. (2021). Effects of device limitations on acquisition of the /t/-/k/ contrast in children with cochlear implants. *Ear & Hearing*, Epub ahead of print, doi: 10.1097/AUD.0000000000001115

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of english

fricatives. *The Journal of the Acoustical Society of America*, *108*(3), 1252–63.

Kant, A. R., Patadia, R., Govale, P., Rangasayee, R., & Kirtane, M. (2012). Acoustic analysis of speech of cochlear implantees and its implications. *Clinical and Experimental Otorhinolaryngology*, *5*, 14–8. https://doi.org/10.3342/ceo.2012.5.S1.S14

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: evidence from acoustic studies. *Journal of Speech and Hearing Research*, *3*, 421-47.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, *73*(1), 322-335.

Kewley-Port, D. & Luce, P. A. (1984) Time-varying features of initial stop consonants in auditory running spectra: A first report. *Attention, Perception and Psychophysics*, *35*(4), 353-360. doi:10.3758/BF0320633

Kim, J., & Chin, S. B. (2008). Fortition and lenition patterns in the acquisition of obstruents by children with cochlear implants. *Clinical Linguistics & Phonetics*, *22*(3), 233–51. https://doi.org/10.1080/02699200701869925

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*(5044), 606-608.

Kumle, L., Võ, M. L., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: an open introduction and tutorial in R. *Behavior Research Methods*, *53*, 2528-2543. doi:10.3758/s13428-021-01546-0

Lee, D. S., Lee, J. S., Oh, S. H., Kim, S. K., Kim, J. W., Chung, J. K., Lee, M. C., & Kim, C. S. (2001). Cross-modal plasticity and cochlear implants. *Nature*, *409*(6817), 149–50.

Lee, S. (2020). Spectral analysis of English fricatives /s/ and /ʃ/ produced by people with profound hearing loss. *Clinical Archives of Communication Disorders*, *5*(3), 147–153. https://doi.org/10.21849/cacd.2020.00269

Li, F., Edwards, J., & Beckman, M. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics*, *37*(1), 111-124.

Liker, M., Mildner, V., & Sindija, B. (2007). Acoustic analysis of the speech of children with cochlear implants: a longitudinal study. *Clinical Linguistics & Phonetics*, *21*(1), 1–11.

Ling, D. (1976). *Speech and the hearing-impaired child: Theory and practice*. Washington, DC: Alexander Graham Bell Association for the Deaf.

Ling, D. (1989). *Foundations of spoken language for the hearing-impaired child*. Washington, DC: Alexander Graham Bell Association for the Deaf.

Loizou, P. (2006). Speech processing in vocoder-centric cochlear implants. *Advances in Otorhinolaryngology, 64*, 109-143.

Macken, M. A., Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word initial stop consonants. *Journal of Child Language*, *7*, 41–74.

Matthies, M. L., Svirsky, M. A., Lane, H. L., & Perkell, J. S. (1994). A preliminary study of the effects of cochlear implants on the production of sibilants. *The Journal of the Acoustical Society of America, 96*(3), 1367–73.

Maxwell, E., & Weismer, G. (1982). The contribution of phonological, acoustic, and perceptual techniques to the characterization of a misarticulating child's voice contrast for stops. *Applied Psycholinguistics, 3*(1), 29-43.

McLeod, S. (2017). *Speech Sound Acquisition*. In J. Bernthal, N. Bankson, & P. Flipsen Jr, Articulation and phonological disorders: Speech sound disorders in children (8th ed., pp. 49-92). Boston, MA: Pearson.

McLeod, S. & Bleile, K. (2003). Neurological and developmental foundations of speech acquisition. *American Speech-Language-Hearing Association Convention*. Chicago, IL.

McMurray, B. & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, *118*(2), 219-246. doi:10.1037/a0022325

Moberly, A. C., Lowenstein, J. H., & Nittrouer, S. (2016). Word recognition variability with cochlear implants: “perceptual attention” versus “auditory sensitivity.” *Ear & Hearing*, *37*(1), 14–26. https://doi.org/10.1097/AUD.0000000000000204

Moore, J. K., & Linthicum, F. H. (2007). The human auditory system: a timeline of development. *International Journal of Audiology*, *46*(9), 460–478.

Munson, B., Edwards, J., Schellinger, S., Beckman, M., & Meyer, M. (2010). Deconstructing phonetic transcription: covert contrast, perceptual bias, and an extraterrestrial view of Vox Humana. *Clinical Linguistics & Phonetics*, *24*(4-5), 245-260. doi:10.3109/02699200903532524

Neumeyer, V., Schiel, F., & Hoole, P. (2015). Speech of cochlear implant patients: An acoustic analysis of sibilant production. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*. Glasgow, Scotland.

Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America*, *109*(3), 1181–96.

Nicholson, N., Reidy, P., Munson, B., Beckman, M. E., & Edwards, J. (2015). The acquisition of English lingual sibilant fricatives in very young children: Effects of age and vocabulary size on transcribed accuracy and acoustic differentiation. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*. Glasgow, Scotland.

Nissen, S. L., & Fox, R. A. (2005). Acoustic and spectral characteristics of young children's fricative productions: a developmental perspective. *The Journal of the Acoustical Society of America*, *118*(4), 2570–8.

Nittrouer, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *Journal of the Acoustical Society of America*, *97*(1), 520-530.

Nittrouer, S., & Lowenstein, J. H. (2015). Weighting of acoustic cues to a manner distinction by children with and without hearing loss. *Journal of Speech, Language, and Hearing Research: JSLHR*, *58*(3), 1077–1092. https://doi.org/10.1044/2015_JSLHR-H-14-0263

Noble, J.H., Labadie, R.F., Gifford, R.H., Dawant, B.M. (2013). Image-guidance enables new

methods for customizing cochlear implant stimulation strategies. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(5).

Noiray, A., Wieling, M., Abakarova, D., Rubertus, E., & Tiede, M. (2019). Back from the future: nonlinear anticipation in adults‘ and children’s speech. *Journal of Speech, Language, and Hearing Research: JSLHR*, *62*(8S), 3033–3054. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0208

Nossair, Z. & Zahorian, S. (1991). Dynamic spectral shape features as acoustic correlates for initial stop consonants. *The Journal of the Acoustical Society of America*, *89*(6), 2978-2991. doi:10.1121/1.400735

Oller, D., Eilers, R., Neal, A., & Schwartz, H. (1999). Precursors to speech in infancy: The prediction of speech and language disorders. *Journal of Communication Disorders, 32*(4), 223-245. doi:10.1016/S0021-9924(99)00013-1.

Osberger, M. (1997). Cochlear implantation in children under the age of two years: Candidacy considerations. *Otolaryngology--Head and Neck Surgery, 117*(3), 145-9.

Osberger, M. & McGarr, N. (1982) *Speech production characteristics of the hearing impaired*. In N. Lass (Ed.), Speech and language: Advances in basic research and practice volume 8 (pp. 227-288). New York: Academic Press.

Peng, Z.E., Hess, C., Saffran, J., Edwards, J., & Litovsky, R. (2019). Assessing fine-grained speech discrimination in young children with bilateral cochlear implants. *Otology & Neurotology, 40*(3), 197. doi:10.1097/MAO.0000000000002115

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Reidy, P. F. (2016) Spectral dynamics of sibilant fricatives are contrastive and language specific. *The Journal of the Acoustical Society of America*, *140*(4), 2518-2529. doi:10.1121/1.4964510

Reidy, P., Kristensen, K., Winn, M., Litovsky, R., & Edwards, J. (2017). The acoustics of word-

initial fricatives and their effect on world-level intelligibility in children with bilateral cochlear implants. *Ear & Hearing, 38,* 42-56.

Romeo, R., Hazan, V., & Pettinato, M. (2013). Developmental and gender-related trends of intra-talker variability in consonant production. *The Journal of the Acoustical Society of America*, *134*(5), 3781–92. https://doi.org/10.1121/1.4824160

Scobbie, J. M., Gibbon, F., Hardcastle, W. J., & Fletcher, P. (2000). *Covert contrast as a stage in the acquisition of phonetics and phonology*. In M. B. Broe & J. B. Pierrehumbert (Eds.), Papers in Laboratory Phonology V: Acquisition and the Lexicon, (pp. 194-207). Cambridge, England: Cambridge UP.

Schatz, T., Feldman, N. H., Goldwater, S., Cao, X. N., & Dupoux, E. (2021). Early phonetic learning without phonetic categories: insights from large-scale simulations on realistic input. *Proceedings of the National Academy of Sciences of the United States of America, 118*(7). https://doi.org/10.1073/pnas.2001844118

Serry, T. A. and Blamey, P. J., (1999). A four-year investigation into phoneme acquisition in young cochlear implant users. *Journal of Speech, Language and Hearing Research: JSLHR*, *42*(1), 141-154.

Skinner, M. W., Fourakis, M. S., Holden, T. A., Holden, L. K., & Demorest, M. E. (1999). Identification of speech by cochlear implant recipients with the multipeak (mpeak) and spectral peak (speak) speech coding strategies ii. Consonants. *Ear & Hearing*, *20*(6), 443.

Smit, A. B., Hand, L., Freilinger, J., Bernthal, J., Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, *55*(4), 779-798.

Smith, L. (1975). Residual hearing and speech production in deaf children. *Journal of Speech and Hearing Research*, *18*, 795-811.

Spencer, L., & Guo, L. (2012). Consonant development in pediatric cochlear implant users who were implanted before 30 months of age. *The Journal of Deaf Studies and Deaf Education, 18*(1), 93-109.

Stevens, K. (1998). Acoustic phonetics. Cambridge: The MIT Press.

Stevens, K.N. & Blumstein, S.E. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, *64*(5), 1358-1368.

Stoel-Gammon, C. (2001). Transcribing the speech of young children. *Topics in Language Disorders, 21*(4), 12-21. doi:10.1097/00011363-200108000-00004

Sundarrajan, M., Tobey, E. A., Nicholas, J., & Geers, A. E. (2020). Assessing consonant production in children with cochlear implants. *Journal of Communication Disorders, 84*. https://doi.org/10.1016/j.jcomdis.2019.105966

Todd A. E., Edwards J. R., & Litovsky R. Y. (2011). Production of contrast between sibilant fricatives by children with cochlear implants. *The Journal of the Acoustical Society of America*, *130*(6), 3969-3979. doi:10.1121/1.3652852

Uchanski, R. M., & Geers, A. E. (2003). Acoustic characteristics of the speech of young cochlear implant users: a comparison with normal-hearing age-mates. *Ear & Hearing*, *24*(1), 90.

Uchanski, R. M., Miller, K. M., Reed, C., M., and Braida, L. D. (1992). *Effects of token variability on vowel identification*. In Schouten (Ed.) The Auditory Processing of Speech: From Sounds to Words. Mouton de Gruyter, Berlin.

Umeda, N. (1977). Consonant duration in American English. *The Journal of the Acoustical Society of America*, *61*(3), 846-858. doi:10.1121/1.381374

Van Lierde, K., Vinck, B., Baudonck, N., De Vel, E., & Dhooge, I. (2005). Comparison of the overall intelligibility, articulation, resonance, and voice characteristics between children using cochlear implants and those using bilateral hearing aids: a pilot study. *International Journal of Audiology, 44*(8), 452–465.

Warner-Czyz, A. D., & Davis, B. L. (2008). The emergence of segmental accuracy in young cochlear implant recipients. *Cochlear Implants International*, *9*(3), 143–166. https://doi.org/10.1002/cii.364

Werker, J., & Tees, R. (2002). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 25*(1), 121-

133. doi:10.1016/S0163-6383(02)00093-0

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

Zeng, F., Rebscher, S., Harrison, W.V., Sun, X., & Feng, H. (2008). Cochlear implants: System design, integration and evaluation. *IEEE Reviews in Biomedical Engineering*, *1*, 115-142. doi:10.1109/RBME.2008.2008250