ABSTRACT

Title of Thesis: INTELLIGIBILITY IN CHILDREN WITH COCHLEAR IMPLANTS: THE /T/ VS. /K/ CONTRAST

Elinora Claire Leonard, Master of Arts, 2019

Thesis Directed By: Professor, Dr. Jan Edwards, Department of Hearing and Speech Sciences

Previous research has found that the speech of children with cochlear implants (CI) is less intelligible than the speech of peers with normal hearing (NH). This claim has been supported by research showing that children with CIs have difficulty with the late-acquired spectral contrast of /s/ vs. /ʃ/: correctly produced words containing these initial-consonants are less intelligible when produced by children with CIs relative to children with NH. The current study examined whether a similar result is observed with the early-acquired spectral contrast of /t/ vs. /k/. Crowd-sourced data were used to evaluate intelligibility of /t/- and /k/-initial words correctly produced by children with CIs and children with NH embedded in multi-talker babble. Results indicated that whole-word productions of children with CIs were less intelligible than productions of children with NH for words beginning with this early-acquired contrast. However, results also indicated this difference in intelligibility was not dependent on the intelligibility of the initial consonant alone.

INTELLIGIBILITY IN CHILDREN WITH COCHLEAR IMPLANTS: THE /T/ VS. /K/ CONTRAST

by

Elinora Leonard

Thesis 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
Master of Arts
2019

Advisory Committee:
Professor Dr. Jan Edwards, Chair
Dr. Rochelle Newman
Dr. Margaritis Fourakis

Graduate Student co-advisor:
Allison Johnson

Preface

While children with cochlear implants (CIs) have much better speech and language skills than children with hearing aids with the same level of hearing loss, their speech and language skills lag behind that of their peers with normal hearing, even after many years of experience with the device (Osberger et al., 1993; Spencer et al., 1998, 1999; Tomblin et al., 1999; Uchanski & Geers et al., 2003). For example, speech intelligibility of children with CIs has been found to be poorer than children with normal hearing even after seven years of cochlear implant use (Peng, Spencer, & Tomblin, 2004). Decreased speech intelligibility can significantly impact the ability of a child to communicate and socialize. The formation of meaningful peer relationships has been shown to lead to higher academic achievement and better integration into adult society (DeLuzio & Girolametto, 2011).

Because the CI processor degrades spectral contrasts (e.g., /t/ vs. /k/) more than temporal (e.g., /t/ vs. /d/) or manner contrasts (e.g., /t/ vs. /s/), place-of-articulation contrasts are particularly difficult for children with CIs (Friesen, Shannon, Baskent, & Wang, 2001; Iverson, 2002; Munson, Donaldson, Allen, Collison, & Nelson, 2002). For example, the contrast between /s/ and /ʃ/ is represented by differences in the concentration of energy in the frication: /s/ has a higher concentration of spectral energy at higher frequencies relative to /ʃ/. Several studies (Todd et al., 2011; Reidy et al., 2017) have suggested that the contrast between /s/ and /ʃ/ is difficult for children with CIs to produce due to limitations in their perception of the sounds. Reidy et al. (2017) found that when children with CIs produced /s/ and /ʃ/ in word-initial position, there was acoustically less contrast between the two sounds compared to productions of children with normal hearing. This result implies that children's ability to produce a contrast is influenced by their ability to perceive spectral differences.

There has been little research on whether children with cochlear implants have similar difficulty with the acquisition of place-of-articulation contrasts that are acquired early. For example, the contrast /s/ and /ʃ/ is typically acquired by 7 years of age. However, /t/ and /k/, another place contrast, is often acquired by 4 years of age (Smit et al., 1990).

The current study addresses the following research question: Are /t/- and/k/-initial words produced by 3-5-year-old children with CIs less intelligible than words produced by their peers with normal hearing? To evaluate this, I will ask adults with normal hearing to identify correctly produced words by children with cochlear implants and children with normal hearing that begin with either /t/ or /k/ in a challenging listening environment (multi-talker babble). Adults' accuracy in identification will be used as the measure of speech intelligibility.

Acknowledgements

I would like to thank my thesis committee, my advisor Dr. Jan Edwards, for all of her support and guidance. I have been fortunate enough to be a part of Dr. Edwards' research team during my senior undergraduate year and throughout graduate school and it has been a wonderful experience. I would also like to thank my other committee members, Dr. Rochelle Newman and Dr. Margaritis Fourakis for their thoughtful feedback throughout this process. I would like to thank Allison Johnson, my graduate student co-advisor, whose assistance and support were paramount. I would also like to thank Zachary Maher for all of his help with data collection through Mechanical Turk and for assistance with analyzing the data. I also want to thank the undergraduate students who helped me with coding stimuli, Elisa Rademacher, Megan Gigliotti, and Michele Liquori. Finally, I would like to thank my husband, parents, and friends for all their support and encouragement throughout this process.

Table of Contents

List of Tables

List of Figures

The process of developing speech sounds throughout infancy and childhood is gradual. As children's anatomy and physiology become more adult-like and as they gain more motor control, they produce a wider variety of sounds more accurately (Kent, 1992). By age 5, children produce the majority of speech sounds correctly (Smit et al., 1990). Sounds that require less articulatory precision to produce are acquired earlier, such as stops, which involve rapid ballistic movements. Stops (e.g., /b/, /p/, /d/, /t/, /k/, /g/) require only a brief occlusion of air by the articulators. In comparison, fricatives (e.g., /s/, /z/, /f/, /v/, /ʃ/, /ʒ/, /θ/, /ð/) require a longer, semi-occluded stream of air and are acquired later in development. In addition, children learn how to coordinate oral and velopharyngeal closure (oral stop vs. nasal) before distinctions of voiced and voiceless (/p/ vs. /b/) and anterior vs. posterior (/k/ vs. /t/) sound articulations (Kent, 1992).

Analyses of children's speech-sound production have shown that children's articulatory gestures are not as well-coordinated as adults' (Nittrouer, 1995). One example that occurs early during typical development is velar fronting (Smit, 1993). Velar fronting occurs when sounds that are produced near the velum are produced closer to the front of the mouth. This substitution can occur for several speech-sound pairs, including /t/ for /k/ and /d/ for /g/. It has been proposed that velar fronting may occur due to poor tongue control: differential anterior and posterior tongue movements (i.e., raising the tongue dorsum while keeping the blade low) are required for velar sounds such as /k/ and /g/ (Heng, 2016). According to the Iowa-Nebraska Articulation Norms project, between the ages of 2 and 4, children produce errors on /k/ about 16% of the time (Smit, 1993). The most common error is to substitute an alveolar stop (i.e., /t/) for a velar stop (i.e., /k/). By ages 4-5, errors on /k/ reduce to 5%, with /t/ for /k/ substitutions remaining as the most common error (Smit, 1993).

Gibbon et al. (1999) described this phenomenon as "undifferentiated lingual gestures." Undifferentiated lingual gestures are produced when the "anterior tongue-palate contact occurs simultaneously with posterior contact" (p. 393). As Gibbon points out, mature lingual control is characterized by the tongue blade and the tongue body having the ability to move independently of each other. Ambiguous productions of /k/ (where different areas of the tongue move as one) often sound more like /t/. As a result, listeners have difficulty classifying sounds when these undifferentiated lingual gestures occur (Gibbon, 1999).

Children who are born deaf or hard of hearing have particular challenges when perceiving and producing speech sounds. Around 90% of children with hearing loss are born to parents with normal hearing (Mitchell, 2004). Assistive listening devices such as hearing aids and cochlear implants are often used in order to help transmit as much sound as possible. CIs are surgically implanted to electrically stimulate the auditory nerve of children with severe to profound sensorineural hearing loss. CIs have grown in popularity over the past 25 years as better outcomes in speech, language, and literacy development have been shown compared to children with severe to profound hearing loss who use hearing aids or no assistive device (Baudonck, 2011; Chin & Finnegan, 2002; Chin et al., 2003; Lejeune & Demanez, 2006).

Baudonck (2011) compared intelligibility in children with CIs to children who use hearing aids (HA). 92% of the 7-year-old children with bilateral CIs in the study were judged as intelligible in daily situations by two speech-language pathologists. Only 40% of the 8-year-old children with HAs in the study were judged as intelligible in daily situations. All participants were fitted with their first HA before the age of 21 months and had at least 18 months of experience with their current devices (HA or CI). Intelligibility judgments were made using a five-point scale, where only a score of 5 is considered intelligible. (The scale was 1 = "totally unintelligible speech," 2 = "nearly unintelligible speech, some single words are intelligible while lipreading and using a known context," 3 = "an intelligible speech if the listener is concentrated and reads the child's lips," 4 = "an intelligible speech for listeners with little experience with deaf speech," and 5 = "an intelligible speech for all listeners in daily situations".)

In another intelligibility study, children with CIs scored higher than children with HAs on an auditory performance scale and an intelligibility scale, even after receiving the same audio-phonatory training (Lejeune & Demanez, 2006). The study included 34 children with severe congenital hearing loss and one with progressive hearing loss. The average age of the HA group was 9 years, and the CI group was 7 years. Auditory development was rated using the Category of Auditory Performance (CAP) score (Archbold, 1995). The CAP score provides an ordinal scale of auditory receptive abilities. In addition, a Speech Intelligibility Rating (SIR) was given as well (Cox & McDaniel, 1989). SIR is a 5-point scale which was designed as a “time-effective global outcome measure of speech production in real-life situations,” (Lejeune & Demanez, 2006, p.65).

Although children with CIs have more intelligible speech than children with HAs, speech sound development for children with CIs is slower than that of children with NH (Baudonck, 2011; Lejeune & Demanez, 2006; Osberger et al., 1993; Spencer et al., 1998, 1999; Tomblin et al., 1999; Uchanski & Geers et al., 2003). In addition, the speech of children with CIs is not as intelligible as that of their peers with normal hearing, which can significantly impact the ability of a child to communicate and socialize (Bat-Chava et al., 2013; DeLuzio & Girolametto, 2011; Martin et al., 2011).

This is not surprising because, compared to acoustic hearing, CIs provide a highly degraded speech signal that contains substantially less information than the normal speech waveform. One issue has to do with frequency resolution. The signal processor in the device breaks up the signal into different frequency bands. For example, channel 1 (the lowest frequency) for the Nucleus 22 (Cochlear™ brand) is set at 250Hz; the bandwidth is 125Hz (188Hz – 313Hz). Channel 22 (the highest frequency) is set at 7,438Hz; the bandwidth is 1000Hz (6938Hz – 7938Hz) (Ali et al., 2015). This demonstrates how much precise frequency information is lost. It also demonstrates, how as frequency increases, the bandwidth of the frequency band increases. Therefore, frequency resolution is poorer at higher frequencies (Ali et al., 2015; Friesen et al., 2001; Munson et al., 2003). For comparison, the human ear can discriminate changes in frequency of signals of about 2%. For example, a change from 100Hz to 102Hz (increase of 2%) can be detected (Seikel et al., 2000, p.609). As previously stated, CIs with 22-channels compress frequencies where one electrode is stimulated for as great a frequency difference as 1000Hz in the high frequency range (Ali et al., 2015).

Spectral contrasts, including place-of-articulation contrasts (e.g., /s/ vs. /ʃ/), have been shown to be more impacted by the poor frequency resolution relative to voicing (e.g., /t/ vs. /d/) and manner contrasts (e.g., /t/ vs. /s/) (Friesen, 2001; Reidy et al., 2017; Rødvik et. al., 2018; Todd et al., 2011;). For example, the voicing contrast (e.g., /p/ vs. /b/) relies on temporal information which the device is better at transmitting. However, spectral contrasts rely more heavily on peak frequency. The mean spectral energy (or "centroid") for /s/ (approximately 8000 Hz) is often above the frequency limitations of the devices, making it more difficult for children with CIs to perceive the contrast between /s/ and /ʃ/ (Todd et al., 2011).

Place-of-articulation contrasts have also been found difficult to produce for CI users, even for post-lingually deafened adults with CIs. One recent systematic review and meta-analysis by, Rødvik et. al., 2018, found the most common errors made on consonants produced by adults who were post-lingually deafened across 50 studies were place-of-articulation contrasts (/k/ as /t/, /m/ as /n/, and /p/ as /t/) (Rødvik et. al., 2018). One specific study, Munson et. al., (2003), gave adults with cochlear implants who were post-lingually deafened speech sound perception tasks with both consonants and vowels. The participants with CIs were split into two groups: those who performed "better" and those who performed "poorer." The groups were defined through total percent-correct identification for consonants and vowels as well as individual listener's scores. There was a clear division between "better" and "poorer" performing groups, as no composite scores fell between 60% and 69%. Specific patterns of phoneme confusion revealed that most errors were due to a place-of-articulation confusion. For example, the ten most common consonant errors for the "better" performing group were t/p, t/k, g/d, f/θ, θ/f, θ/s, v/ð, ð/v, ð/z, and m/n (where t/k means the stimulus /k/ was misperceived as /t/). The ten most common errors for the "poorer" performing adults almost entirely overlapped, with the exception of additional difficulty with k/p and r/l. These results demonstrate that the groups' overall performance did not differ qualitatively between their perception of consonants (Munson, 2003). Although CI devices provide some spectral information to the user, perception and production abilities are limited, especially with regard to place-of-articulation.

Place-of-articulation contrasts can be even more difficult for children with CIs who are prelingually deafened. Place-of-articulation errors, as well as other phonological errors, can significantly affect overall intelligibility. Phonological acquisition of children with CIs is, ideally, closely monitored throughout early childhood. There are many ways of measuring single word productions for both clinical and research purposes. One way is through standardized articulation tests such as the *Goldman-Fristoe Test of Articulation-Third Edition* (GFTA-3; Goldman & Fristoe, 2015), the *Manchester Junior Word List* (Watson, 1957), or short phrase repetitions such as the *Beginners' Intelligibility Test* (Osberger, Robbins, Todd, & Riley, 1994) and *Bamford-Kowal-Bench sentences* (Bamford, 1979) (e.g., Chin, 2003; Connor et al., 2000; Dornan, 2009; Geers et al., 2000;). Listener judgements of speech productions have also been used to measure speech intelligibility (Baudonck, 2011; Lejeune et al., 2006; Reidy et al., 2017). Intelligibility is often measured by ratings on intelligibility scales such as the *Speech Intelligibility Rating* (SIR; Cox,

1989) (e.g., Baudonck, 2011; Lejeune et al., 2006) or having the listeners repeat back what they heard the speaker say (e.g., Reidy et al., 2017; Todd et al., 2011).

Chin, Tsai, and Gao (2003) evaluated speech intelligibility in English-speaking children with and without cochlear implants. They found that speech intelligibility of children with cochlear implants increased with both age and longer device use. In addition, the *Beginners' Intelligibility Test* was used to compare intelligibility between groups. Listeners judged the connected speech of 2-10-year-old children with CIs to be 35% accurate, compared to 87% accurate for 2-6-year-old children with NH.

Reidy et al. (2017) specifically examined the /s/-/ʃ/contrast produced by children with CIs and by age-matched peers with normal hearing. There were two primary questions. The first was whether there were spectral differences between correct productions of /s/- and /ʃ/-initial words in both groups, and the second was whether there were differences in intelligibility between groups. The researchers found that spectral features differed between groups: children with normal hearing produced /s/ with a higher peak frequency--and thus produced a more robust /s/-/ʃ/ contrast--compared to children with CIs.

To determine whether spectral differences in production impacted intelligibility, adult participants listened to correct productions of /s/- and /ʃ/-initial words in noise and were recorded repeating what words they heard. The percent accuracy of the adults’ responses was used as the measure of intelligibility. Results showed that productions by children with CIs and NH were equally intelligible overall; however, the /s/-initial words produced by children with CIs were less intelligible than their /ʃ/-initial words. The lower peak frequency of /s/ productions by the children with CIs impacted intelligibility, suggesting that sub-phonemic differences in productions have consequences for intelligibility.

Reidy et al. (2017) proposed that the reduced contrast observed between these sound pairs is due to degraded signals provided by the CI. The signal contains poor spectral resolution which makes it difficult for children to acquire certain place-of-articulation contrasts. Reidy et al. provides crucial evidence that children with CIs produce sounds with different spectral features than their peers with normal hearing, and when words are intelligible in quiet environments, they may be unintelligible in more adverse listening conditions.

While a number of studies have examined acquisition of the /s/-/ʃ/ contrast, there has been little research on whether children with cochlear implants have similar difficulty with earlier-acquired place contrasts, such as /t/-/k/. The /t/-/k/ contrast is usually acquired by age 4, but may pose particular difficulty for children with CIs, because they are short, transient sounds and also differentiated by degraded spectral cues.

The purpose of this study is to measure the intelligibility of /t/- and /k/-initial words produced by children with cochlear implants and children with normal hearing in a challenging listening environment.

This challenging listening environment was created by using multi-talker babble. Multi-talker babble balances the effects of energetic masking and informational masking in order to reflect everyday listening environments. Acoustic characteristics of the babble overlap with acoustic characteristics of the target speech which can lead to masking at the periphery (energetic masking) and can make target speech difficult to extract from the babble (informational masking). Four-talker babble has been used in previous studies (e.g., Reidy et al., 2017), because it reduces informational masking relative to two-talker babble but does not eliminate it completely which occurs as the number of talkers increases. The ability to decode individual words in babble decreases as the number of talkers is further increased which could lead to similar effects than that of white noise.

This study differs from previous studies in several ways. First, word-level and initial-consonant specific intelligibility of the early-acquired /t/-/k/ contrast has not been studied as extensively as the later acquired /s/-/ʃ/ contrast. Second, to emulate real-world listening scenarios, this study will examine intelligibility of only words that have been produced correctly in a challenging listening environment.

Methods

Participants

Speakers:

Stimuli for this study were isolated words produced by 26 children with cochlear implants between the ages of 31 to 66 months (mean age of 50 months) and 26 children with normal hearing matched for age, sex, and maternal education. The mean age of implantation for the children with cochlear implants was 17 months. Recordings were made in a sound-treated lab setting at either the University of Wisconsin-Madison or the University of Minnesota-Twin Cities. All participants produced target words during a picture-prompted, auditory word-repetition task. See Table 1 for demographic information on both groups of children.

Table 1.

Demographic information for the two speaker groups.

Group	Age (months)	Male:Female	Maternal Education
CI	50 range: 31-66	11:15	High school or less: 2 Some college/2-year degree: 5 College or graduate degree: 19
NH	50 range: 31-66	11:15	High school or less: 2 Some college/2-year degree: 5 College or graduate degree: 19

Listeners:

I recruited adults to participate in a listening task through Amazon Mechanical Turk (Mturk), which is a crowdsourcing website. Of approximately 1000 listeners, the results of 683 participants were analyzed. All of these 683 participants self-reported as monolingual, native speakers of American English and were between the ages of 18 to 40. All participants also had at least a 90% acceptance rating on Mturk, meaning their work was rarely rejected. Each listener was randomly assigned to listen to a given list of words. In a similar in-lab task by Reidy et al. (2017), 4 adult listeners were used per speaker. McAllister et al. (2014) examined the validity of Mechanical Turk ratings and found that 9 Mturk workers "Mturkers" were needed to achieve the same validity as one in-lab listener. Therefore, the goal was to recruit 36 listeners per child:(4 in-lab listeners (as in Reidy et al., 2017) x 9 Mturk listeners (as in McAllister et al., 2014). Each listener heard only one production of each target word. The participants were compensated through Amazon Mechanical Turk. The lists were broken down into payment "bins" depending on how many tokens were in each list, as follows: 1.) 14-17 tokens ($0.22); 2.) 18-20 tokens ($0.25); 3.) 20-21 tokens ($0.27); and 4.) 22-24 tokens ($0.31). This worked out to approximately 0.13 cents per token.

Materials

Speech Stimuli:

Target words included 15 different /t/-initial words (e.g., "tummy") and 17 different /k/-initial words (e.g., "kitty") produced by children with CIs and children with NH, as described in Table 1 above. Words were recorded as part of a larger longitudinal study at the University of Wisconsin-Madison and the University of Minnesota-Twin Cities. There were also filler words that began with other consonants (/w/, /r/, /s/, and /ʃ/). Words were familiar to children based on age-of-acquisition norms, picturable (words were easily represented by a supporting image), and balanced across four vowel contexts (high/mid-front (/i, ɪ, e/), mid/low-front (/ɛ, æ/), high/mid-back (u, ʊ, o/), mid/low-back (/ɑ, ɔ, ʌ/)). Some words were repeated in order to balance consonant/vowel contexts within a particular age range. The words are provided below in Table 2. During the task, a picture stimulus appeared on a computer screen in front of the child while the corresponding word was presented over speakers. Children repeated the word into a microphone, and responses were recorded.

Stimulus Selection

All initial consonants had already been transcribed as correct or incorrect for a previous study (Johnson, Bentley, Munson, & Edwards, 2019). If the initial consonant was transcribed as correctly produced, then the entire word was listened to by me or an undergraduate student trained in phonetic transcription. Words were judged as being produced entirely correctly or containing at least one phonetic error. Word productions were included only if they contained no errors. Only one repetition of each target word within a speaker was included. All repetition decisions were made based on which of the two productions was produced with more clarity (i.e., least amount of background noise, least amount of clipping, etc.) All words were amplitude-normalized at 74 dB. On average, the children with CIs produced 17 words correctly, and the children with NH produced 18 words correctly.

Table 2.

Words contained in each wordlist. Wordlist 1 was used for age range 31- to 39-month-olds, wordlist 2 for age range 40- to 51-month-olds, and wordlist 3 for age range 52- to 66-month-olds.

Word	IPA transcription	Vowel Context	Wordlist: repetitions
Cake	/ke͜ɪk/	High/mid-front	1: 2 2: 1 3: 1
Candle	/kændl̩/	Mid/low-front	1: 0 2: 1 3: 1
Candy	/kændi/	Mid/low-front	1: 2 2: 1 3: 2
Car	/kɑɹ/	Mid/low-back	1: 2 2: 1 3: 0
Cat	/kæt/	Mid/low-front	1: 2 2: 1 3: 1
Catch	/kæt͜ʃ/	Mid/low-front	1: 0 2: 1 3: 0
Coat	/ko͜ʊt/	High/mid-back	1: 2 2: 1 3: 1
Coffee	/kɑfi/	Mid/low-back	1: 0 2: 1 3: 1
Comb	/ko͜ʊm/	High/mid-back	1: 0 2: 1 3: 1

Cookie	/kʊki/	High/mid-back	1: 2 2: 2 3: 2
Cousin	/kʌzɪn/	Mid/low-back	1: 0 2: 0 3: 1
Cup	/kʌp/	Mid/low-back	1: 2 2: 1 3: 1
Cutting	/kʌɾɪŋ/	Mid/low-back	1: 0 2: 1 3: 1
Keys	/kiz/	High/mid-front	1: 0 2: 1 3: 1
Kitchen	/kɪt͜ʃɪn/	High/mid-front	1: 2 2: 1 3: 1
Kitten	/kɪtn̩/	High/mid-front	1: 0 2: 0 3: 1
Kitty	/kɪɾi/	High/mid-front	1: 2 2: 1 3: 0
Table	/te͜ɪbl̩/	High/mid-front	1: 2 2: 1 3: 1
Take	/te͜ɪk/	High/mid-front	1: 0 2: 1 3: 1
Tape	/te͜ɪp/	High/mid-front	1: 2 2: 1 3: 0
Teacher	/tit͜ʃɹ̩/	High/mid-front	1: 0 2: 0 3: 1

Teddy bear	/tɛdibe͜ɪɹ/	Mid/low-front	1: 2 2: 2 3: 2
Tent	/tɛnt/	Mid/low-front	1: 0 2: 2 3: 2
Tickle	/tɪkl̩/	High/mid-front	1: 2 2: 1 3: 1
Tiger	/ta͜ɪgɹ̩/	Mid/low-back	1: 0 2: 1 3: 1
Toast	/to͜ʊst/	High/mid-back	1: 2 2: 1 3: 1
Toaster	/to͜ʊstɹ̩/	High/mid-back	1: 0 2: 1 3: 2
Tongue	/tʌŋ/	Mid/low-back	1: 2 2: 2 3: 1
Tooth	/tuθ/	High/mid-back	1: 2 2: 1 3: 0
Toothbrush	/tuθbɹʌʃ/	High/mid-back	1: 0 2: 1 3: 1
Towel	/taʊl̩/	Mid/low-back	1: 0 2: 0 3: 1
Tummy	/tʌmi/	Mid/low-back	1: 2 2: 1 3: 2

Multi-talker Babble:

The multi-talker babble used for the present study consisted of speech produced by four adult female talkers taken from several corpora. One speaker was taken from the IEEE corpus (IEEE, 1969), another from the BKB corpus (Bench et al., 1979), and two from the AzBio corpus (Spahr et al., 2012). These speakers were chosen because they have been used in previous work (e.g., Reidy et al., 2017).

Multi-talker babble was added 1 second before, throughout, and 1 second after each target word. The babble was set with an 8 dB signal-to-noise (SNR) ratio after running a short pilot study. The pilot study included five lists of tokens from NH participants, each presented with the following five SNRs: 0dB, 3dB, 6dB, 8dB, and 10dB. Results from 500 listeners (20 per SNR, per list) showed that approximately 70% accuracy was achieved with an SNR of 8dB. Thus, an 8dB SNR was selected for this study.

Data Collection

Listening Task:

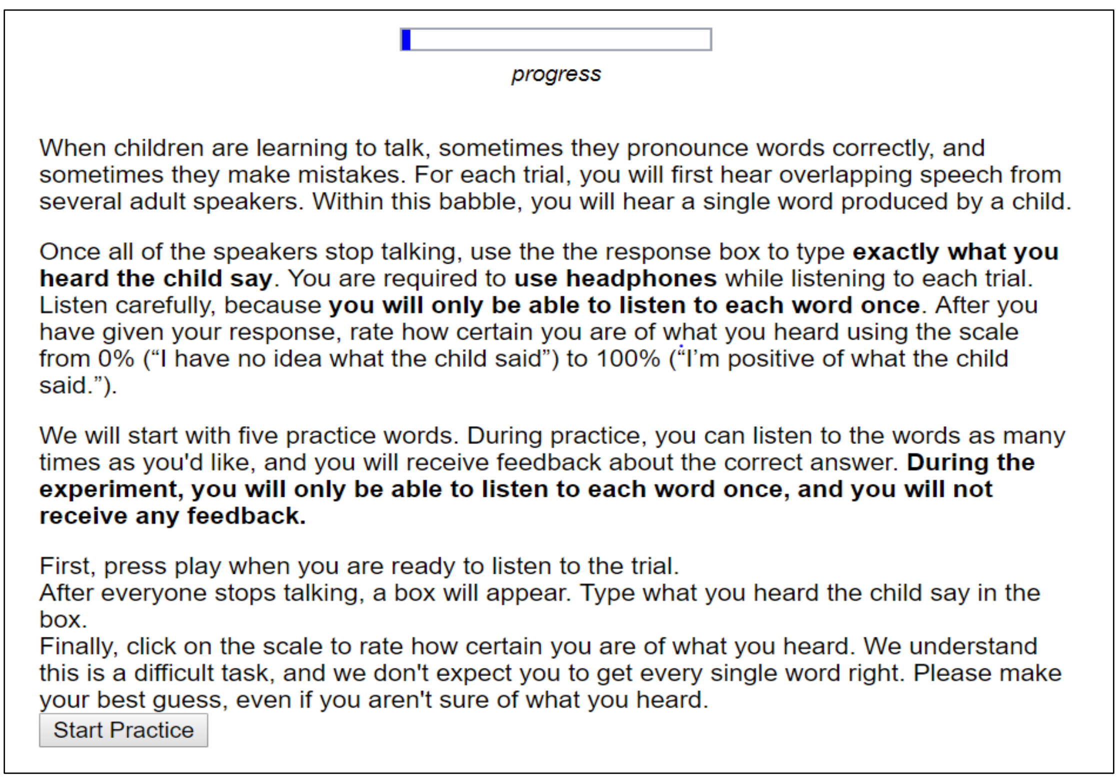
The experiment was created using Ibex PennController (Zehr & Schwarz, 2018), and the link to the experiment was provided on Amazon Mechanical Turk. The participants were asked to listen to single-word productions from a single child in babble and fill in a text-entry box with what they heard. The participants were also asked to rate how certain they were of what they heard on a continuum (0% "I have no idea what the child said," to 100% "I'm positive of what the child said"). The lists were randomly assigned (1 list per listener), and the number of tokens per list ranged from 14 to 24. There was a minimum of 4 /t/-initial words and 4 /k/-initial words on each list. The total number of filler words varied across lists with an average of 9 filler words per list (range = 2 to 14).

Mturkers could select to participate if they met the initial criteria (90% acceptance rating on Mturk, located in the United States). Participants were given a brief study description:

> “Listen to a variety of short sound clips. In each sound clip, you will hear overlapping speech from several adult speakers. Within this babble, you will hear a word produced by a young child. Your task is to type the word you hear.”

If Mturkers chose to participate after reading the study description, they were first asked to read and “sign” the consent form. After they consented, they were taken to an instructions page. See Figure 1.

Figure 1. Listener Instructions.



progress

When children are learning to talk, sometimes they pronounce words correctly, and sometimes they make mistakes. For each trial, you will first hear overlapping speech from several adult speakers. Within this babble, you will hear a single word produced by a child.

Once all of the speakers stop talking, use the the response box to type **exactly what you heard the child say**. You are required to **use headphones** while listening to each trial. Listen carefully, because **you will only be able to listen to each word once**. After you have given your response, rate how certain you are of what you heard using the scale from 0% (“I have no idea what the child said”) to 100% (“I’m positive of what the child said.”).

We will start with five practice words. During practice, you can listen to the words as many times as you'd like, and you will receive feedback about the correct answer. **During the experiment, you will only be able to listen to each word once, and you will not receive any feedback.**

First, press play when you are ready to listen to the trial.
After everyone stops talking, a box will appear. Type what you heard the child say in the box.
Finally, click on the scale to rate how certain you are of what you heard. We understand this is a difficult task, and we don't expect you to get every single word right. Please make your best guess, even if you aren't sure of what you heard.

Start Practice

Then the participant was asked to confirm they were wearing headphones. See Figure 2.

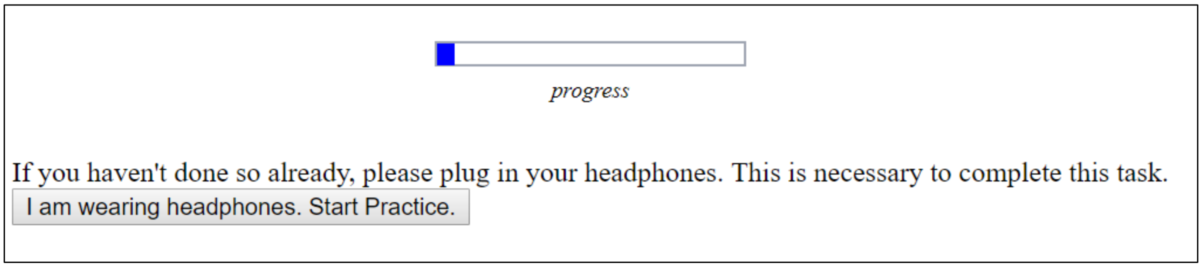


Figure 2. Headphones Reminder

Each participant was then given 5 practice words in which they could listen to each word multiple times, and they were told the target word at the end. See Figure 3.

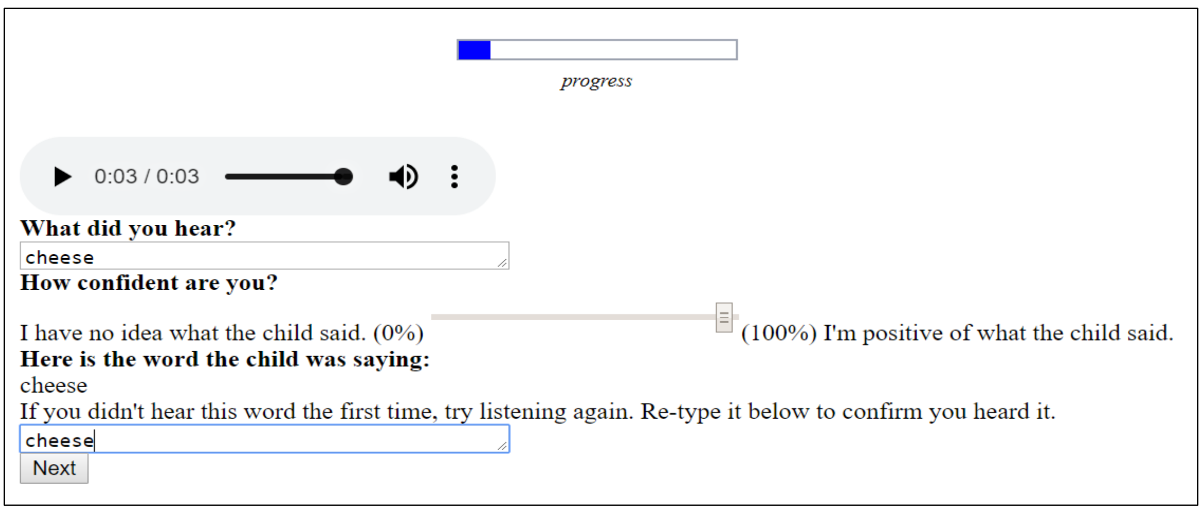


Figure 3. Practice Trial Example.

After the practice words, the participants were told the study was going to begin and were reminded that they would only hear each audio file once. See Figure 4.

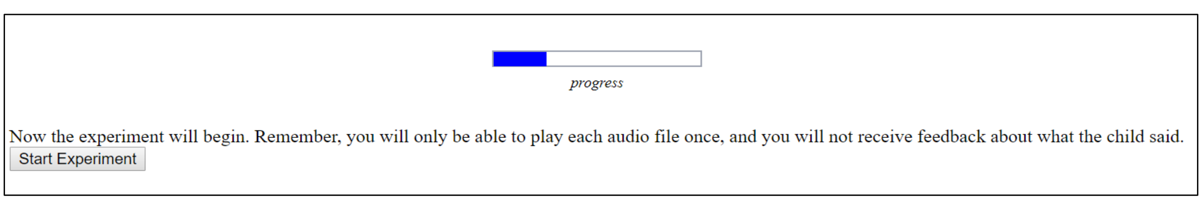


Figure 4. Start experiment.

Each trial of the study required a fill-in-the-box response and a confidence rating before the option to continue appeared. See Figure 5.

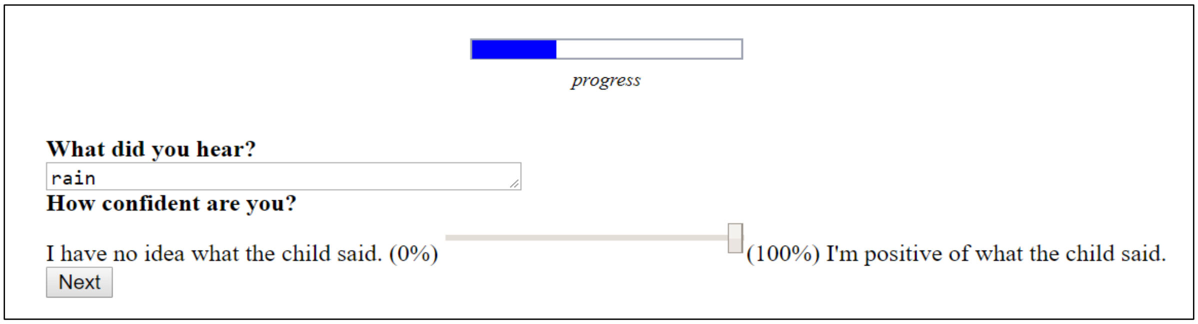


Figure 5. Trial Example.

Included among the target words, were four “catch words,” where the SNR was set at 15dB. These words were used to gauge the participant’s effort level during the task. A minimum of 75% accuracy on the catch words was required for the participant’s responses to be included in the analysis.

Coding:

For the whole-word accuracy coding, the listener’s typed response had to match the target word in order to be scored as correct (1). If the listener’s response did not match the target, it was scored as incorrect (0). For the initial-consonant accuracy coding, the listener’s typed response for the initial-consonant had to match the target consonant in order to be scored as correct (1). More specifically, /k/ was coded as correct if the response started with “k” or “c” and the second letter was not “h,” or the consonant-vowel sequences “ci” or “ce.” /t/ was coded as correct if the response started with “t” and the second letter was not “h.”

Data Analysis:

Participant data were excluded if they were outside of the acceptable age range, reported a language other than English as their primary language, completed the task more than once, or responded inaccurately on more than 1 of the 4 catch words.

In order to determine whether there are group differences in intelligibility, I used a mixed-effects logistic regression model. A logistic regression model was appropriate because the dependent variable was binary (the response either matched the target or did not). A mixed- effects model was necessary because of the multiple sources of non-independence in the data (i.e., multiple tokens per child, multiple responses per word). The model predicted accuracy based on fixed effects of Group (CI vs. NH) and Target Consonant (/t/ vs. /k/) as well as by-speaker and by-target-word random intercepts. There were two homologous mixed-effects logistic regression models used to analyze the effects of front and back vowel contexts separately. A mixed-effects model was also used to predict confidence ratings based on fixed effects of Group (CI vs. NH), Target Consonant (/t/ vs. /k/), Accuracy (correctly identified vs. incorrectly identified) as well as by-speaker and by-target-word random intercepts. Finally, a mixed-effects logistic regression model was used to analyze the accuracy results of the initial-consonant only. The model predicted accuracy based on the fixed-effects of Group (CI vs. NH) and Target Consonant (/t/ vs. /k/) as well as by-speaker and by-target-word random intercepts. Then, as with the whole-word accuracy analysis, there were two homologous mixed-effects logistic regression models used to analyze the effects of front and back vowel contexts separately.

Predictions:

Based on previous research, I made two predictions. First, I predicted there would be a significant effect of Group, such that the correct productions of children with CIs would be less intelligible than productions of their age-matched peers, for both the whole-word analysis and for the initial consonant analysis. Second, I predicted that confidence ratings would be lower for productions of children with CIs relative to productions of children with NH.

Results

Ideally, there would be 36 ratings per token. Results reported here reflect an average of 7.91 ratings per token for the children with CIs and 4.27 ratings per token for the children with NH. Obtaining 36 ratings per token was not feasible within the time frame for this thesis, due to diminishing response rates on Mturk (only Mturkers located in the United States with a 90% acceptance rate were allowed to participate, and each Mturker could only rate one list), as well as the exclusionary criteria applied to some participants who responded (listeners needed to be 18-40 years old and report English as their primary language). Other contributing factors were the low compensation rate (1.3 cents per token on average) and the short list length (also contributing to the low compensation rate). Table 3 provides detailed information on the number of tokens rated for each speaker. Table 4. Provides information on how the speakers' productions were broken down by session or time point (different ages), number of lists, and mean length of lists.

Table 3.

Number of tokens rated for each speaker

Speaker Group	Speaker ID	Number of tokens	Average number of ratings	Median number of ratings	Minimum number of ratings	Maximum number of ratings
CI	302E	13	22.00	22	22	22
CI	303E	27	4.04	5	3	5
CI	304E	25	9.42	7	3	23
CI	306E	13	21.46	27	15	27
CI	308E	13	3.00	3	3	3
CI	309E	23	4.04	4	3	5
CI	311E	19	4.32	4	4	5
CI	314E	20	7.27	7	7	8
CI	605L	21	7.48	7	7	8
CI	608L	29	3.96	4	1	9
CI	665L	21	11.92	9	5	21
CI	679L	28	11.21	5	3	27
CI	801E	26	4.00	4	2	5
CI	804E	23	3.74	4	2	5
CI	807E	16	2.00	2	2	2
CI	809E	21	6.33	6	5	8
NH	002L	31	3.73	4	2	5
NH	010L	22	3.45	3	3	4
NH	014L	18	2.13	3	1	3
NH	030L	27	5.14	4	2	8
NH	037L	15	7.67	8	6	9
NH	050L	11	3.00	3	3	3
NH	052L	30	2.64	3	1	4
NH	053L	27	3.93	3	2	7
NH	057L	8	7.00	7	7	7
NH	058L	21	2.57	4	1	4
NH	063L	30	4.59	5	2	8
NH	072L	23	6.22	7	4	9
NH	076L	9	4.00	4	4	4
NH	078L	25	2.57	3	1	3
NH	099L	26	5.74	5	4	8
NH	100L	23	3.17	2	2	5
NH	126L	22	4.27	4	4	5
NH	615L	24	2.96	2.5	1	6

Table 4.

Number of sessions/time points, number of lists, and mean length of lists

Speaker ID	Session ID (Time point)	Number of Lists	Mean length of lists
002L	002L51FS4	3	20
002L	002L63FS6	4	16
010L	010L44MS3	2	20
014L	014L39MS2	1	17
014L	014L50MS4	2	18.5
030L	030L38FS2	2	14.5
030L	030L49FS4	4	16.75
037L	037L62MS6	4	15.25
050L	050L36FS2	1	16
052L	052L43MS3	4	16
052L	052L55MS5	3	17.67
053L	053L47MS4	3	15.67
053L	053L59MS6	3	19
057L	057L58MS6	2	15.5
058L	058L48FS4	2	19
063L	063L47FS4	4	15.25
063L	063L59FS6	3	19.67
072L	072L55FS5	3	15
076L	076L57MS6	3	17.33
078L	078L51FS4	3	17.33
078L	078L63FS6	3	18.33
099L	099L51FS4	2	14.5
099L	099L64FS6	2	19.5
100L	100L63MS6	4	16.5
126L	126L51MS5	3	20
302E	302E49FS2	2	12.5
303E	303E65FS2	4	17
304E	304E48FS2	2	16.5
304E	304E59FS3	3	13.67
306E	306E64FS3	2	12
308E	308E37FS1	1	15
309E	309E59MS2	3	16.67
311E	311E62MS2	3	16
314E	314E38FS1	1	14
314E	314E50FS2	2	14
605L	605L43MS3	2	15.5

605L	605L55MS5	3	14.67
608L	608L39FS2	2	14
608L	608L52FS4	3	20
608L	608L64FS6	3	19
612L	612L55FS5	4	17
615L	615L66FS6	4	17
640L	640L60FS6	4	17
665L	665L52FS4	2	10.5
665L	665L64FS6	3	14
679L	679L46MS4	2	12
679L	679L58MS6	3	18
801E	801E38MS1	1	20
801E	801E50MS2	3	19.67
804E	804E56MS2	3	19
807E	807E51MS2	3	17.67
809E	809E64MS2	3	14.67

Whole-word Accuracy

The primary research question was whether children with cochlear implants were less intelligible than their peers with normal hearing. We assessed this question by running a logistic mixed-effects regression model with fixed-effects of Group (CI vs. NH), Target Consonant (/t/ vs./k/), and the interaction between Group and Target Consonant, as well as random intercepts for Speaker and Target Word. The dependent variable was Accuracy on the perception task. The final model was coded in *R* (Rstudio Team, 2016) using the lmerTest package (Kuznetsova et al., 2017) as follows: (Accuracy ~ Group + Target Consonant + Group * Target Consonant + (1|Speaker ID) + (1|Target Word)). The reference conditions were children with CIs (Group) and the consonant /t/ (Target Consonant). Model summaries for this model and all subsequent models are provided in the Appendix.

The by-Speaker random intercept was found to significantly improve model fit ($x_1^2 = 180, p =< 0.001$); as was the by-Target Word random intercept ($x_1^2 =$

$600, p =< 0.001$); indicating that there was significant variability in both of these factors. The effect of Target Word is shown below in Figure 6. It can be observed that, for both groups, some words were identified more accurately than others.

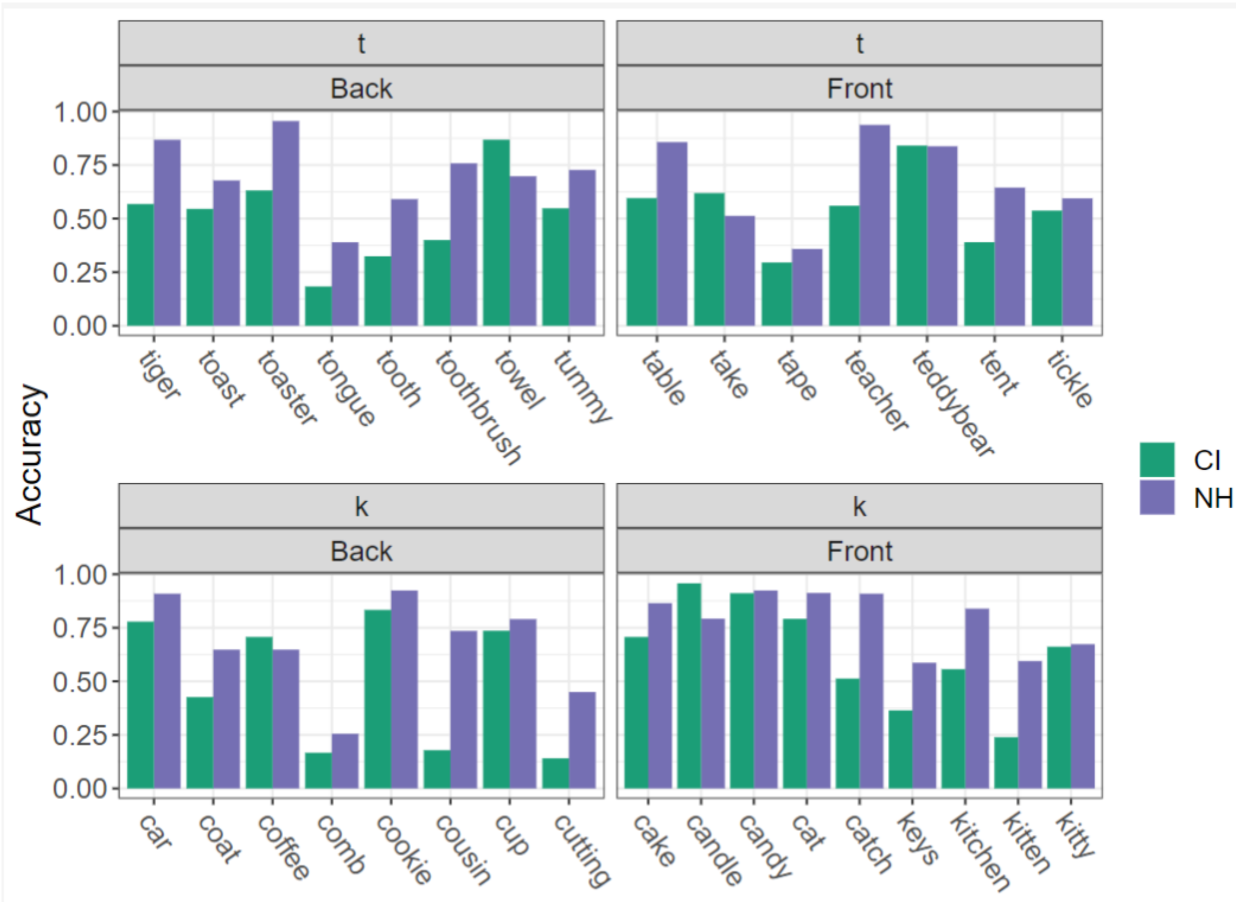


Figure 6. Mean accuracy by Word and Group, for each Target Consonant and Vowel Context.

Overall, there was a significant main effect of Group ($\hat{\beta}$ = 0.57, *SE* = 0.21, *z* = 2.72, *p* = 0.00647). Listeners were more accurate in identifying the words produced by children with NH compared to words produced by children with CIs. This significant main effect is illustrated in Figure 7. There was no main effect of Target Consonant (β = 0.19, *SE* = 0.34, *z* = 0.56, *p* = 0.57367). Listeners were equally accurate in identifying /t/-initial words and /k/-initial words. There was also no significant interaction between Group and Target Consonant (β = 0.09, *SE* = 0.14, *z* = 0.62, *p* = 0.53888). The difference in accuracy between productions by children with CIs and those by children with NH were similar across consonants.

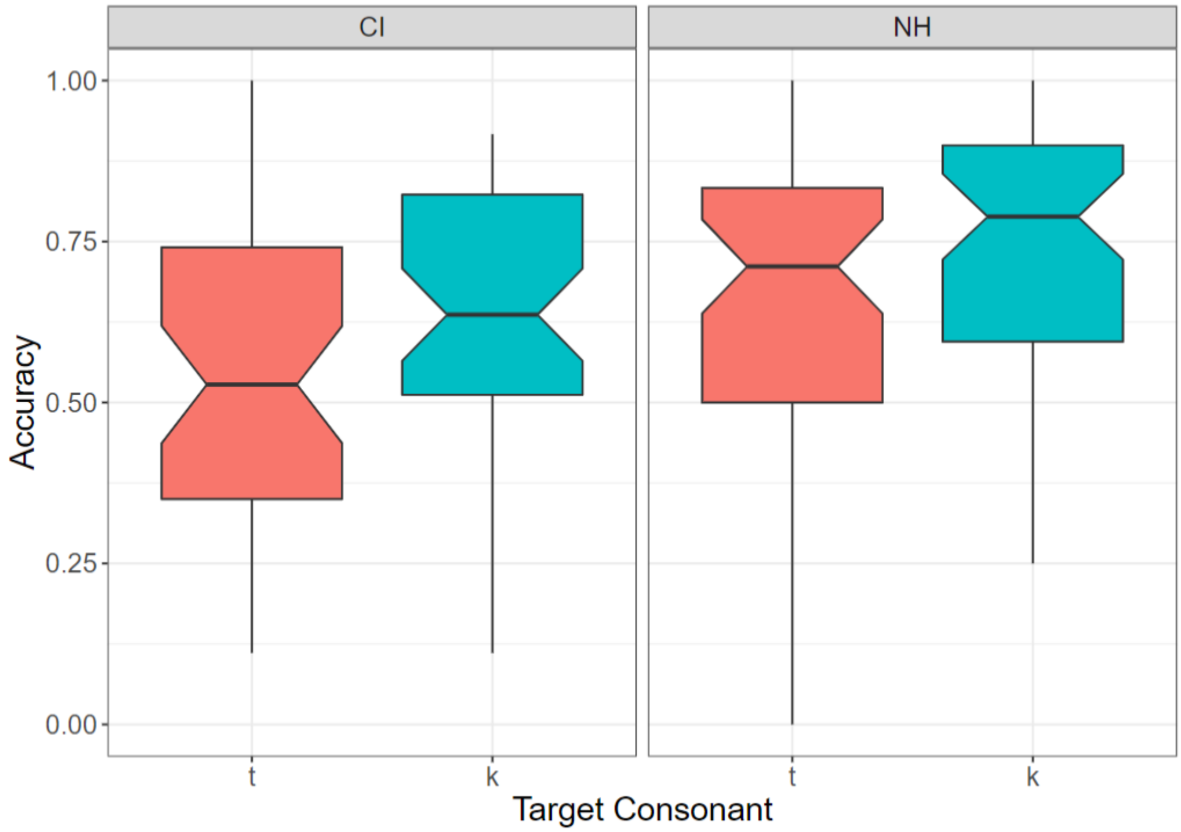


Figure 7. Whole-word accuracy by Group and Target Consonant. Horizontal line indicates median, boxes indicate inter-quartile range, and whiskers indicate minimum and maximum values

I also ran two homologous logistic mixed-effects regression models, one for productions in front vowel contexts and one for productions in back vowel contexts. Figure 8 shows accuracy separately for the two vowel contexts. The same *R* code and reference conditions were used for these two models. In the front vowel context, there were no significant main effects.

In the back vowel context, there was a significant main effect of Group (β = 0.77, *SE* = 0.25, *z* = 3.08, *p* = 0.0021); listeners were more accurate at identifying the words produced by children with NH compared to children with CIs. There was no main effect of Target Consonant (β = 0.03, *SE* = 0.52, *z* = 0.06, *p* = 0.9522) and no significant interaction between Group and Target Consonant (β = -0.22, *SE* = 0.21, *z* = -1.03, *p* = 0.3050), indicating that listeners' accuracy in identifying /k/-initial and /t/-initial words was comparable across groups.

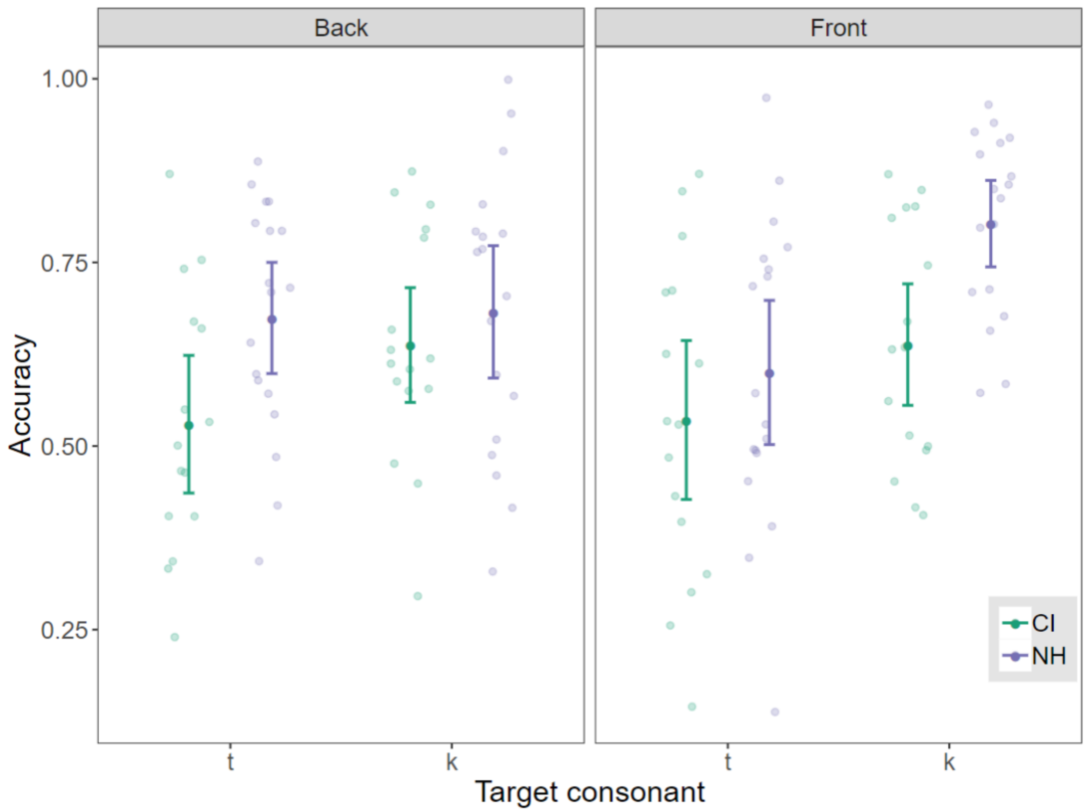


Figure 8. Accuracy by Group and Target Consonant, plotted separately for front and back vowel contexts. Circles indicate individual data points; solid lines show mean and +/- two standard error.

Confidence Ratings

For the continuous rating scale, the dependent variable was the confidence ratings, converted into a percentage ranging from 0 to 100%. Fixed effects included Group, Target Consonant, Accuracy (whether or not the word was accurately identified), all 2-way interactions, and the 3-way interaction. There were also random intercepts for Speaker and Target Word. The reference conditions were children with CIs (Group), the consonant /t/ (Target Consonant), and 0 for the Accuracy condition (i.e., not correct). The final model was coded in *R* (Rstudio Team, 2016) using the lmerTest package (Kuznetsova et al., 2017) as follows: Confidence Ratings ~ Group * Target Consonant * Accuracy + (1|Speaker ID) + (1|Target Word).

All main effects and interactions were significant. The main effect of Group ($\beta = 6.75$, $SE = 2.34$, $t = 18.11$, $p = < 0.001$) indicated that for target consonant /t/ when the word was intelligible, confidence ratings were higher for children with NH compared to children with CIs. The main effect of Target Consonant ($\beta = 9.73$, $SE = 3.08$, $t = 3.15$, $p = 0.003$) indicated that for children with CIs when the word was not identified accurately, confidence ratings were higher for words that start with /k/ compared to /t/. The main effect of Accuracy ($\beta = 36.94$, $SE = 1.52$, $t = 24.22$, $p < 0.001$) indicated that for children with CIs, words beginning with /t/ that were identified accurately had higher confidence ratings than words that were not identified accurately. The interaction between Group and Target Consonant ($\beta = -9.26$, $SE = 2.67$, $t = -3.47$, $p < 0.001$) indicated that for children with NH, the difference in ratings between the two consonants /t/ and /k/ was smaller than the difference in ratings between the consonants for children with CIs. These main effects and the Group by Target Consonant interaction are illustrated in Figure 9. The interaction between Group and Accuracy (β = -6.20, *SE* = 2.38, *t* = -2.60, *p* = 0.009) indicated that for children with NH, the relationship between confidence and accuracy was weaker than for children with CIs. The interaction between Target and Accuracy (β = 7.10, *SE* = 2.04, *t* = -3.48, $p < 0.001$) indicated that for words that start with /k/, the relationship between ratings and accuracy was weaker than for words that start with /t/. The 3-way interaction (β = 7.71, *SE* = 3.29, *t* = 2.35, *p* = 0.02) indicated that the relationship between Group and Accuracy differed between consonants. For children with CIs, there was a greater difference in confidence ratings between the two consonants in the lower accuracy range and listeners were more confident on productions with initial /k/ than productions with initial /t/. For children with NH, there was also a greater difference in confidence ratings between the two consonants in the lower accuracy range, but listeners were more confident on productions with initial /t/ than productions with initial /k/. The Group by Accuracy interaction, the Target Consonant by Accuracy interaction and the three-way interaction are illustrated in Figure 10.

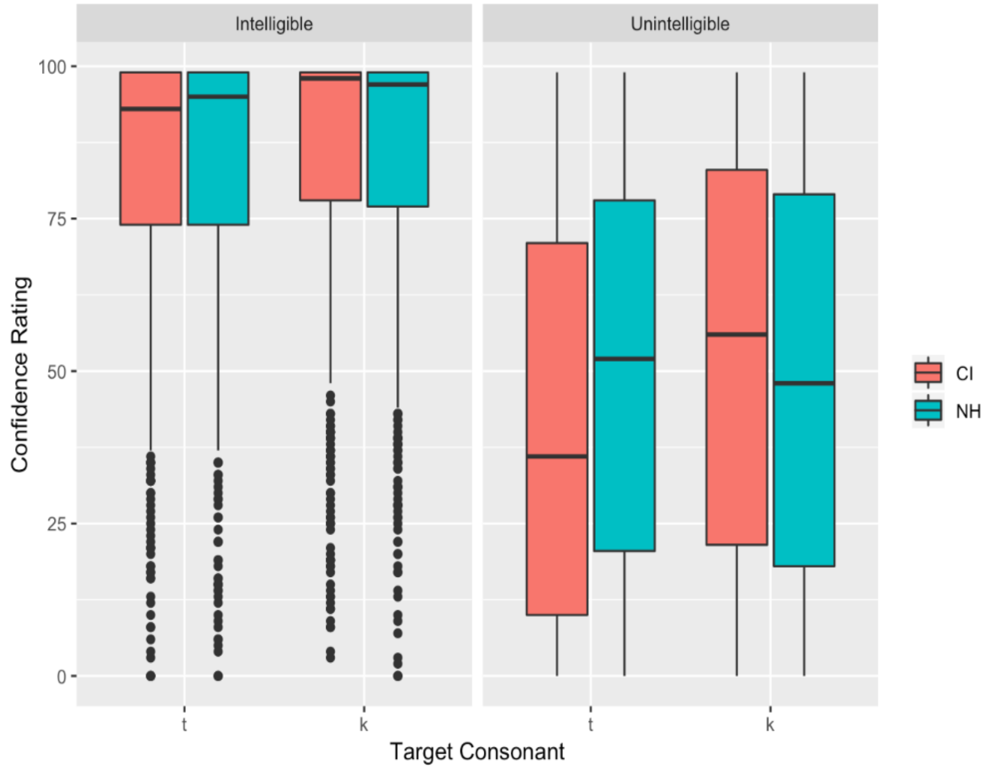


Figure 9. Ratings by Target Consonant for words identified correctly and words not identified correctly for children with CIs and children with NH. Horizontal line indicates median, boxes indicate inter-quartile range and whiskers indicate minimum and maximum values.

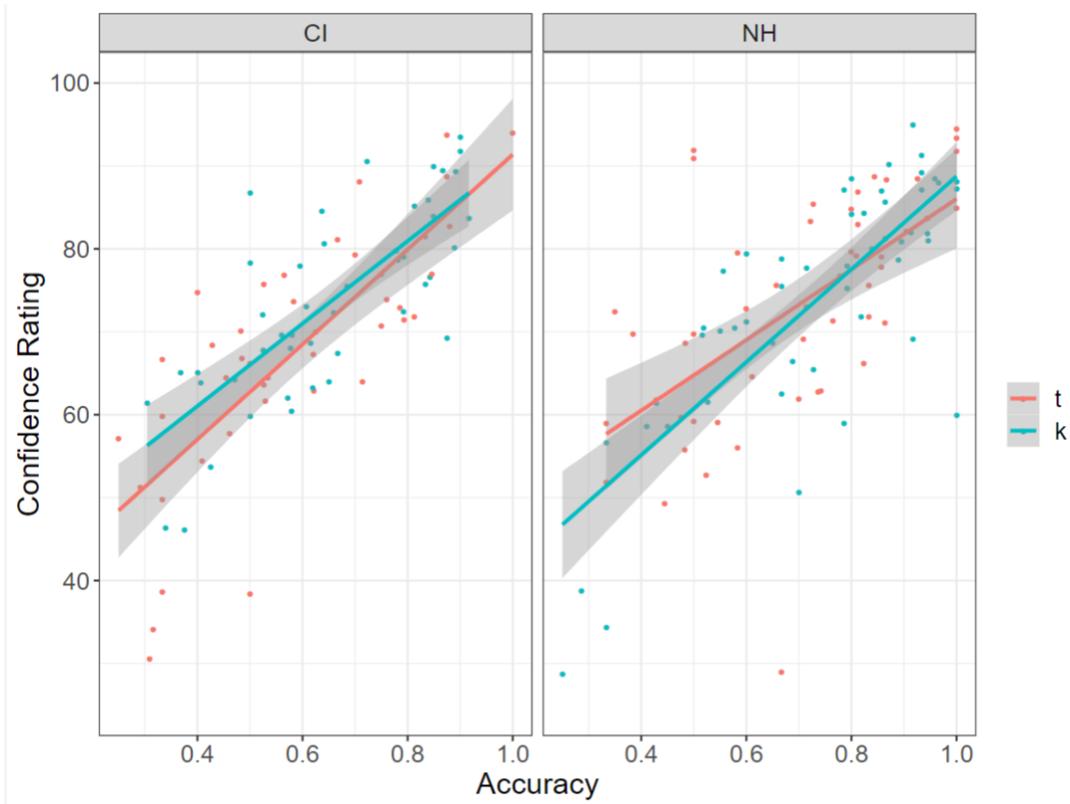


Figure 10. Ratio of words correctly identified (from 0 to 1) plotted against confidence ratings by speaker for /t/ and /k/, plotted separately by group.

<u>*Initial-Consonant Accuracy*</u>

I also ran three models that focused on accuracy of the initial /t/ or /k/ only. For the overall analysis that combined across front and back vowel contexts (see Figure 11), there was no significant main effect of Group (β = .40, *SE* = 0.22, z = 1.78, p = 0.07436); listeners were equally accurate in identifying the initial-consonants produced by children with and the initial-consonants produced by children with CIs. However, the effect of Group was trending towards the initial-consonant productions of children with CIs being found less accurate relative to the initial-consonant productions of children with NH. There was a significant main effect of

Target Consonant (β = 0.78, *SE* = 0.23, *z* = 3.34, *p* = 0.0083); listeners were more accurate in identifying /k/ than /t/. This significant main effect is illustrated in Figure 11. There was no significant interaction between Group and Target Consonant (β = -0.07, *SE* = 0.15, *z* = -0.47, *p* = 0.64055); the difference in accuracy between productions by children with CIs and those by children with NH were similar across consonants.

Two homologous mixed-effects logistic regression models were run separately for front and back-vowel contexts. Figure 12 shows initial-consonant accuracy separately for the two vowel contexts. In the front vowel context, there was no significant main effect of Group (β = .31, *SE* = 0.27, *z* = 1.15, *p* = 0.250); listeners were equally accurate in identifying the initial-consonants produced by children with NH compared to the initial-consonants produced by children with CIs. There was a significant main effect of Target Consonant (β = 0.81, *SE* = 0.36, *z* = 2.26, *p* = 0.024); listeners were more accurate in identifying /k/ than /t/. There was no significant interaction between Group and Target Consonant (β = .18, *SE* = 0.20, *z* = 0.88, *p* = 0.377) This means that the difference in accuracy between productions by children with CIs and those by children with NH were similar across consonants.

In the back-vowel context, the effects and interaction were the same as in the front vowel context. There was no significant main effect of Group (β = .50, *SE* = 0.27, *z* = 1.87, *p* = 0.0616). There was a significant main effect of Target Consonant (β = .81, *SE* = 0.29, *z* = 2.83, *p* = 0.0046), with accuracy being higher for /k/ compared to /t/. There was also not a significant interaction between Group and Target Consonant (β = -0.39, *SE* = 0.22, *z* = 1.77, *p* = 0.077). However, the effect of

Group and the interaction between Group and Target Consonant trended towards significance. There was a trending effect of the initial-consonant productions of children with CIs being found less accurate relative to the initial-consonant productions of children with NH. The trending interaction between Group and Target Consonant suggested that the magnitude of the difference in accuracy between children with CIs and children with NHs was larger for /t/ than /k/ in the back vowel context.

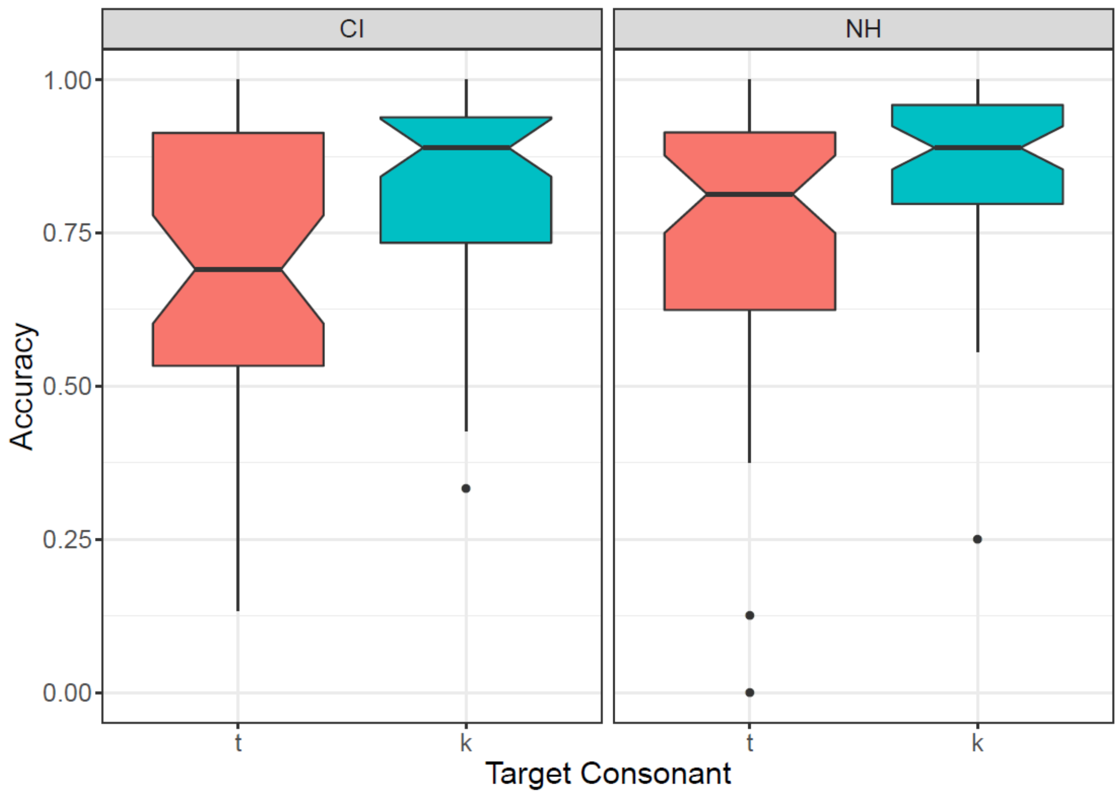


Figure 11. Initial-Consonant Accuracy by Group and Target Consonant. Horizontal line indicates median, boxes indicate inter-quartile range, and whiskers indicate minimum and maximum values.

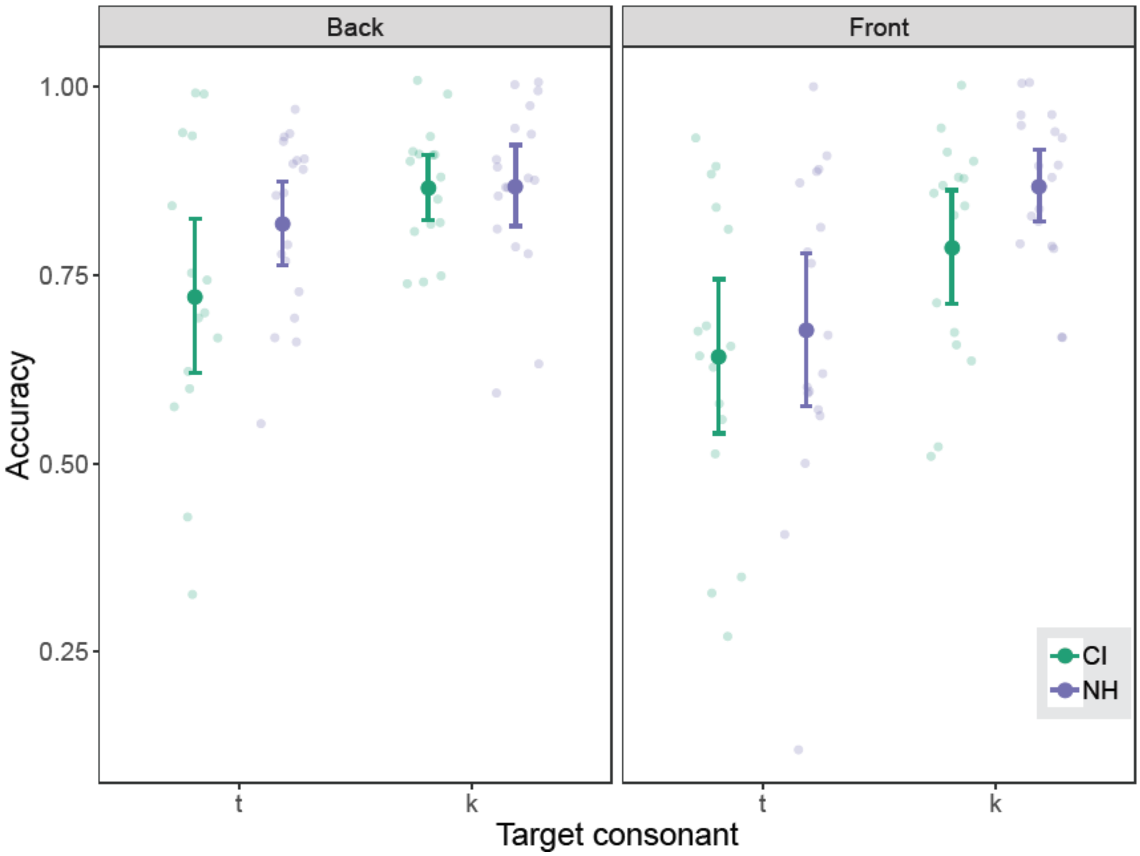


Figure 12. Initial consonant by Group and Target Consonant, plotted separately for front and back vowel contexts. Circles indicate individual data points; solid lines show mean and +/- two standard error.

Discussion

The first prediction I made was only partially supported by the results and the second prediction I made was fully supported, although more complicated than I had anticipated. Overall, whole-word productions of /t/- and /k/-initial words by children with CIs were found to be less intelligible in a challenging listening environment (multi-talker babble) compared to whole-words produced by children with NH. This was observed despite the inclusion of only correct productions. However, this result was not due to less accurate identification of /t/ and /k/ for children with CIs relative to children with NH. When I examined accuracy for initial consonants only, I found that the initial consonants produced by children with CIs were equally intelligible to the initial consonants produced by children with NH. It should be noted that, while Reidy et al. (2017) found a significant difference in intelligibility for a late-acquired spectral contrast (/s/ - /ʃ/), that study did not conduct a similar analysis of initial sounds. Thus, it is unclear whether the lower intelligibility of /s/- and /ʃ/-initial words in Reidy et al. is due to children with CIs having difficulty producing this late-acquired spectral contrast or whether, as in the current study, it is due to less intelligible word productions more generally for children with CIs.

The relationship between confidence ratings and Group was more complex than I had predicted. There was a stronger relationship between accuracy and confidence ratings for the productions from children with CIs, relative to those from children with NH. This finding, in conjunction with the result of the whole-word accuracy analysis, suggests that listeners find it more difficult to understand children with CIs, relative to their peers with NH.

When only the accuracy on initial consonants was analyzed, there was no longer a main effect of Group in the overall or vowel context models, suggesting that the errors seen in the whole-word accuracy analysis likely involved errors in other parts of the word. However, there was a significant main effect of Target Consonant in all models, listeners had more difficultly identifying the initial consonant /t/ compared to the initial consonant /k/, across groups. The consonant /k/ has a stronger burst than the consonant /t/, which may explain why /t/ was more difficult to identify correctly in multi-speaker babble (Repp et. al., 1989). It was unclear why there were no group differences in accuracy for the word-initial consonants. However, accuracy rates were high for both groups when analyzing only the initial consonant, which made subtle differences difficult to observe. The current results suggest that difficulty in producing a contrast between /t/ and /k/ cannot account for the differences in overall word intelligibility between the two groups.

As for the early-acquired t-k- contrast in general, it seems the listeners were not less accurate at identifying initial /k/ and /k/ productions for children with CIs’s relative to those of children with NH. This would suggest that this early-acquired contrast is equally intelligible when produced correctly. However, during the process of stimuli selection, there were more errors on initial /t/ and /k/ for children with CIs, relative to children with NH. This higher accuracy rate of /t/ and /k/ for children with NH relative to children with CIs supports the claim that spectral contrasts are challenging in perception and production for children with CIs.

These results have clinical implications for speech-language pathologists who work on speech production with children with CIs. Primarily, it’s important to continue working on sounds even after they are judged as “correct,” as accuracy alone does not guarantee intelligibility. It’s also important to work on speech production in less ideal environments than a quiet therapy room. For a school speech-language pathologist, this might include working on speech production in the classroom. One reason this is important is that a speech pathologist might think a client with a CI is 100% intelligible, but if this is only in an ideal listening environment, this is not case. Results of this study showed that children with normal hearing have overall higher intelligibility in a challenging listening environment than children with CIs. In addition, a speech pathologist could offer clients with CIs strategies for compensating for reduced intelligibility in noisy settings, such as speaking louder or articulating with more effort.

There were several limitations with the current design. First, as noted above, I was not able to collect 36 ratings per token because of the slow response rate over Amazon Mechanical Turk. Furthermore, it is not clear that Amazon Mechanical Turk is the ideal method of data collection for this type of research. While crowdsourced data over Amazon Mechanical Turk has been successfully used in previous studies on children's speech production (e.g., McAllister et al., 2014), this was a particularly challenging task, which may have contributed to the variability in listener responses that was observed. There may have also been an impact of dialect mismatch due to the speakers being from the Midwest and the listeners having been from many different areas across the United States. One example is vowel variation (i.e., /ketʃ/ vs. /kaetʃ/). Finally, I also did not evaluate accuracy in quiet, which would have provided a useful comparison.

Further analysis of the data could be examine whether errors patterns that listeners made differed by group. Another additional analysis would be to add age to the model. The age range for the speakers included children as young as 31 months and we wouldn't expect to have mastered the /t/-/k/ contrast. Adding chronological age, as well as hearing age, for the children with CIs, could help to tease apart which errors may have simply been developmental in nature.

Both the by-Speaker and by-Target Word random intercepts significantly improved model fit. Future studies could examine potential individual differences between speakers that might be predictive of intelligibility. These could include overall articulation ability, vocabulary size, and hearing age. Future studies could also examine potential differences between words that might be predictive of intelligibility. These could include syllable length, word frequency, and the presence or absence of minimal pairs. It can be observed in Figure 6 that many of the target words with the lowest identification scores had minimal pairs (“cake” vs. “take,” “tape” vs. “cape,” “toast” vs. “coast,” etc.). Minimal pairs can lead to confusion of one word for another because of one phonetic difference. One specific example is “tape” which had one of the lowest accuracy ratings and has the following minimal pairs: “cape,” “gape,” “shape,” among others, versus “teddybear,” which had one of the highest accuracy ratings and has no minimal pairs. Figure 6 also shows that two-syllable words tended to be identified more accurately than one-syllable words. Along with this point, 2-syllable words also tend to have fewer neighbors which can reduce the chance they will be confused with another word.

Regardless of these limitations, this study adds to a growing body of literature indicating that children with CIs produce less intelligible speech than their peers with NH, especially in less than ideal listening environments. These results suggest that speech-language pathologists should consider overall intelligibility as well as transcribed accuracy when they are working with children with CIs and should not be quick to dismiss these children from therapy simply because they can correctly produce targeted sounds in quiet environments.

Appendix

Model summaries

Appendix Table 1.

Overall model of whole-word accuracy scores.

R code: Accuracy ~ Group + Target Consonant + Group * Target Consonant + (1|SpeakerID) + (1|Target Word)

	Estimate	Standard error	*z*-value	*p*-value
(Intercept)	0.23	0.28	0.82	0.41
Group	0.57	0.21	2.72	0.01
Target Consonant	0.19	0.34	0.56	0.57
Group x Target Consonant	0.09	0.14	0.61	0.54

Appendix Table 2.

Model of whole-word accuracy scores in front vowel contexts.

R code: Accuracy ~ Group + Target Consonant + Group * Target Consonant + (1|SpeakerID) + (1|Target Word)

	Estimate	Standard error	*z*-value	*p*-value
(Intercept)	0.23	0.37	0.62	0.53
Group	0.44	0.26	1.67	0.096
Target Consonant	0.37	0.45	0.83	0.41
Group x Target Consonant	0.34	0.21	1.65	0.099

Appendix Table 3.

Model of whole-word accuracy scores in back vowel contexts.

<u>R code</u>: Accuracy ~ Group + Target Consonant + Group * Target Consonant + (1|SpeakerID) + (1|Target Word)

	Estimate	Standard error	*z*-value	*p*-value
(Intercept)	0.19	0.40	0.47	0.64
Group	0.77	0.25	3.08	0.002
Target Consonant	0.03	0.52	0.06	0.95
Group x Target Consonant	-0.22	0.201	-1.03	0.31

Appendix Table 4.

Model of confidence ratings.

R code: Confidence Ratings ~ Group * Target Consonant * Accuracy +

(1|SpeakerID) + (1|Target Word)

	Estimate	Standard error	df	*t*-value	*p*-value
(Intercept)	44.83	2.48	56.70	18.10	< 0.001
Group	6.75	2.34	136.66	2.88	0.005
Target Consonant	9.7	3.08	42.42	3.16	0.003
Accuracy	36.94	1.52	4836.95	24.22	< 0.001
Group x Target Consonant	-9.26	2.67	4827.53	-3.47	< 0.001
Group x Accuracy	-6.12	2.38	4816.75	-2.60	0.009
Target Consonant x Accuracy	-7.10	2.04	4828.03	-3.48	< 0.001
Group x Target Consonant x Accuracy	7.71	3.29	4816.07	2.35	0.019

Appendix Table 5.

Overall model of Initial-consonant accuracy scores.

R code: Accuracy ~ Group + Target Consonant + Group * Target Consonant + (1|SpeakerID) + (1|Target Word)

	Estimate	Standard error	*z*-value	*p*-value
(Intercept)	0.95	0.22	4.27	1.9000
Group	0.40	0.22	1.78	0.07436
Target Consonant	0.78	0.23	3.34	0.00083
Group x Target Consonant	-0.07	0.15	-0.47	0.64055

Appendix Table 6.

Model of initial-consonant accuracy scores in front vowel contexts.

R code: Accuracy ~ Group + Target Consonant + Group * Target Consonant + (1|SpeakerID) + (1|Target Word)

	Estimate	Standard error	*z*-value	*p*-value
(Intercept)	0.77	0.31	2.47	0.013
Group	0.31	0.27	1.15	0.250
Target Consonant	0.81	0.38	2.26	0.024
Group x Target Consonant	0.18	0.20	0.88	0.377

Appendix Table 7.

Model of initial-consonant accuracy scores in back vowel contexts.

R code: Accuracy ~ Group + Target Consonant + Group * Target Consonant +

(1|SpeakerID) + (1|Target Word)

	Estimate	Standard error	*z*-value	*p*-value
(Intercept)	1.15	0.26	4.43	9.4000
Group	0.50	0.27	1.87	0.0616
Target Consonant	0.81	.29	2.83	0.0046
Group x Target Consonant	-0.39	0.22	-1.77	0.0765

References

Ali, H., Noble, J. H., Gifford, R. H., Labadie, R. F., Dawant, B. M., Hansen, J. H., & Tobey, E. (2015). Image-guided customization of frequency-place mapping in cochlear implants.

In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 5843-5847). IEEE.

Archbold, S., Lutman, M., & Marshall, D., (1995). Categories of Auditory Performance. *Ann Otol Rhinol Laryngol Suppl*, *166*, 312–314.

Bat-Chava, Y., Martin, D., & Imperatore, L. (2014). Long-term improvements in oral communication skills and quality of peer relations in children with cochlear implants: Parental testimony. *Child: Care, Health and Development, 40*(6), 870-881. doi:10.1111/cch.12102

Baudonck, N., Van Lierde, K., 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,* 912-919.

Carhart, R., Tillman, T., & Greetis, E. (1969). Perceptual masking in multiple sound backgrounds. *The Journal of the Acoustical Society of America*, *45*(3), 694-703.

Chin, S., & Finnegan, K. (2002). Consonant cluster production by pediatric users of cochlear implants. *Volta Review*, *102,* 157-174.

Chin, S. B., Tsai, P. L., & Gao, S. (2003). Connected speech intelligibility of children with cochlear implants and children with normal hearing. *American Journal of*

Speech–Language Pathology, 12, 440–451.

Ching, T., Psarros, C., Incerti, P., & Hill, M. (2003). Management of children using cochlear implants and hearing aids. *Volta Review, 103,* 39-58.

Connor, C., Craig, H., Raudenbush, S., Heavner, K., & Zwolan, T. (2006). The age at which young deaf children receive cochlear implants and their vocabulary and speech-production growth: Is there an added value for early implantation? *Ear and Hearing, 27*(6), 628-44.

Cox R., & McDaniel M. (1989). Development of the Speech Intelligibility Rating (SIR) Test for Hearing Aid Comparisons. *Journal of Speech and Hearing Research, 32,*347-352

Freyman, R., Balakrishnan, U., & Helfer, K. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America, 115*(5), 2246-56.

Friesen, L., Shannon, R., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America, 110*(2), 1150-63.

Gibbon, F. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research: Jslhr,42*(2), 382.

Habib, M., et al., (2010). Speech production intelligibility of early implanted pediatric cochlear implant users. *International Journal of Pediatric Otorhinolaryngology, 74*(8), 855-9. doi:10.1016/j.ijporl.2010.04.009

Heng, Q., McCabe, P., Clarke, J., & Preston, J. (2016). Using ultrasound visual feedback to remediate velar fronting in preschool children: A pilot study. *Clinical Linguistics & Phonetics, 30*(3-5), 382-97. doi:10.3109/02699206.2015.1120345

Johnson, A., Munson, B., Bentley, D., and Edwards, J. (2019). What can speech acquisition data tell us about cochlear implant device limitations? Analyzing accuracy, error patterns, and spectral features of children's /t/ and /k/ productions. Poster [accepted to be] presented at the 2019 Conference on Implantable Auditory Prostheses. Lake Tahoe, CA, 14-19 July.

Kosky, C., & Boothroyd, A. (2003). Perception and production of sibilants by children with hearing loss: A training study. *Volta Review, 103*, 71-98.

Kuznetsova, A, Brockhoff, PB, Christensen, RHB. (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software*, *82*(13), 1–26. doi: 10.18637/jss.v082.i13.

Lejeune, B. & Demanez L. (2006). Speech discrimination and intelligibility: Outcome of deaf children fitted with hearing aids or cochlear implants. *B-ENT*, *2*(2), 63-8.

Martin, D., Bat-Chava, Y., Lalwani, A., & Waltzman, S. (2011). Peer relationships of deaf children with cochlear implants: Predictors of peer entry and peer interaction success. *Journal of Deaf Studies and Deaf Education, 16*(1), 108-20. doi:10.1093/deafed/enq037

McAllister Byun, T., Halpin, P., & Szeredi, D. (2015). Online crowdsourcing for efficient rating of speech: A validation study. *Journal of Communication Disorders*, *53*, 70-83. doi:10.1016/j.jcomdis.2014.11.003

McCormack, J., Harrison, L., McLeod, S., & McAllister, L. (2011). A nationally representative study of the association between communication impairment at 4-5 years and children's Life activities at 7-9 years. *Journal of Speech, Language, and Hearing Research: Jslhr,54*(5), 1328-48. doi:10.1044/1092-4388(2011/10-0155).

McGowan, R., Nittrouer, S., & Chenausky, K. (2008). Speech production in 12-month-old children with and without hearing loss. *Journal of Speech, Language, and Hearing Research: Jslhr,51*(4), 879-88. doi:10.1044/1092-4388(2008/064).

Mitchell, R., Karchmer, M. (2004). Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the United States. *Sign Language Studies, 4*(2):138-163.

Munson, B., Donaldson, S., Allen, L., Collison, A., & Nelson, A. (2003). Patterns of phoneme perception errors by listeners with cochlear implants as a function of overall speech perception ability. *The Journal of the Acoustical Society of America*, *113*(2), 925-935.

Nittrouer, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *The Journal of the Acoustical Society of America*, *97*(1), 520-30.

Osberger, M., Maso, M., & Sam, L. (1993). Speech intelligibility of children with cochlear implants, tactile aids, or hearing aids. *Journal of Speech and Hearing Research, 36*(1), 186-203.

Peng, S., Spencer, L., & Tomblin, J. (2004). Speech intelligibility of pediatric cochlear implant recipients with 7 years of device experience. *Journal of Speech, Language, and Hearing Research: Jslhr,47*(6), 1227-36.

Phillips, L., Hassanzadeh, S., Kosaner, J., Martin, J., & Deibl, M. (2009). Comparing auditory perception and speech production outcomes: Non-language specific assessment of auditory perception and speech production in children with cochlear implants. *Cochlear Implants Int., 10*(2), 92-102

Pratt, S. (2005). Aural habilitation update: The role of speech production skills of infants and children with hearing loss. *Asha Leader, 10*(4), 8-8. doi:10.1044/leader.FTR2.10042005.8

Reidy, P., et al., (2017). The Acoustics of word-initial fricatives and their effect on word-level intelligibility in children with bilateral cochlear implants. *Ear and Hearing, 38*(1), 42–56. doi:10.1097/AUD.0000000000000349

Repp, H., & Lin, B. (1989). Acoustic properties and perception of stop consonant release transients. *The Journal of the Acoustical Society of America*, *85*(1), 379-396.

Richter, B. (2002). Receptive and expressive language skills of 106 children with a minimum of 2 years' experience in hearing with a cochlear implant. *International Journal of Pediatric Otorhinolaryngology, 64*(2), 111-25.

Rødvik, K., von Koss Torkildsen, J., Wie, B., Storaker, A., & Silvola, T. (2018).

Consonant and Vowel Identification in Cochlear Implant Users Measured by Nonsense Words: A Systematic Review and Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *61*(4), 1023-1050.

RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Sander, E. K. (1972). When are speech sounds learned? *Journal of speech and hearing disorders, 37*(1), 55-63.

Seikel, A., King, D., & Drumright, D. (2000). *Anatomy and Physiology for Speech, Language, and Hearing* (2nd ed.). San Diego, California: Singular Publishing Group.

Smit, A., Hand, L., Freilinger, J., Bernthal, J., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech Language and Hearing Research, 34*(2), 446-446. doi:10.1044/jshr.3402.44

Smit, A. (1993). Phonologic error distributions in the Iowa-Nebraska articulation norms project: Consonant singletons. *Journal of Speech Language and Hearing Research, 36*(3), 533-533. doi:10.1044/jshr.3603.533

Spencer, L., Tye-Murray, N., & Tomblin, J. (1998). The production of English inflectional morphology, speech production and listening performance in children with cochlear implants. *Ear and Hearing, 19*(4), 310-8.

Svirsky, M., Chin, S., Miyamoto, R., Sloan, R., & Caldwell, M. (2002). Speech intelligibility of profoundly deaf pediatric hearing aid users. *Volta Review, 102*, 175-198.

Todd, A., Edwards, J., & Litovsky, R. (2011). Production of contrast between sibilant fricatives by children with cochlear implants. *The Journal of the Acoustical Society of America*, *130*(6), 3969-79. doi:10.1121/1.3652852

Tomblin, J. B., Peng, S., Spencer, L. J., & Lu, N. (2008). Long-term trajectories of the development of speech sound production in pediatric cochlear implant recipients. *Journal of Speech, Language, and Hearing Research*, *51*, 1353-1368.

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

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX)