**[This is a preprint of a manuscript that has been published at JSLHR.**

**The final version is available at https://doi.org/10.1044/2022_JSLHR-21-00209]**

# Recognizing voices through a cochlear implant: A systematic review of voice perception, talker discrimination, and talker identification

Sarah Colby[1], Adriel John Orena[2]

[1]Department of Psychological and Brain Sciences, University of Iowa,
G60 Psychological and Brain Sciences Building, 340 Iowa Ave., Iowa City, Iowa, 52242, USA
[2]Department of Psychology, University of British Columbia,
2136 West Mall, Vancouver, British Columbia, V6T 1Z4, Canada
sarah-colby@uiowa.edu, aorena@psych.ubc.ca

## ABSTRACT

**Objective:** Some cochlear implant (CI) users report having difficulty accessing indexical information in the speech signal, presumably due to limitations in the transmission of fine spectrotemporal cues. The purpose of this paper was to systematically review and evaluate the existing research on talker processing in CI users. Specifically, we reviewed the performance of CI users in three types of talker- and voice-related tasks. We also examined the different factors (such as participant, hearing and device characteristics) that might influence performance in these specific tasks.

**Design:** We completed a systematic search of the literature with select keywords using citation aggregation software to search Google Scholar. We included primary reports that tested: i) talker discrimination; ii) voice perception, and iii) talker identification. Each report must have had at least one group of participants with cochlear implants. Each included study was also evaluated for quality of evidence.

**Results:** The searches resulted in 1561 references, which were first screened for inclusion and then evaluated in full. Forty-three studies examining talker discrimination, voice perception, and talker identification were included in the final review. Most studies were focused on postlingually deafened and implanted adult CI users, with fewer studies focused on prelingual implant users. In general, CI users performed above chance in these tasks. When there was a difference between groups, CI users performed less accurately than their normal-hearing (NH) peers. A subset of CI users reached the same level of performance as NH participants exposed to noise-vocoded stimuli. Some studies found that CI users and NH participants relied on different cues for talker perception. Within groups of CI users, there is moderate evidence for a bimodal benefit for talker processing, and there are mixed findings about the effects of hearing experience.

**Conclusion:** The current review highlights the challenges faced by CI users in tracking and recognizing voices and how they adapt to it. While large variability exists, there is evidence that CI users can process indexical information from speech, though with less accuracy than their NH peers. Recent work has described some of the factors that might ease the challenges of talker processing in CI users. We conclude by suggesting some future avenues of research to optimize real-world speech outcomes.

***Keywords:*** Cochlear implants, Talker perception, Talker identification, Systematic review

## 1. Introduction

Perceiving the indexical properties of the speech signal is a fundamental communicative ability that is often taken for granted (see Sidtis & Kreiman, 2012 for a review). Indeed, many studies have shown that normal-hearing (NH) listeners are adept at perceiving indexical information: they achieve high degrees of accuracy in identifying an unfamiliar talker's age, gender, and their regional background (Bradlow & Bent, 2008; Perry et al., 2001; G. E. Peterson & Barney, 1952). Even infants are able to recognize their mother's voice from birth (DeCasper & Fifer, 1980). Given the intersensory redundancy between faces and voices, it is tempting to dismiss voice recognition as a trivial skill. Indeed, in daily conversations, we can often tell who is speaking because we have visual confirmation. However, familiarity with a speaker's voice does not only allow us to track who is speaking, it also lends itself to efficient social communication and linguistic processing.

For example, familiarity with a talker's voice allows listeners to deal with talker-specific variability in speech. In quiet speech, the same acoustic sound could be distinguished as one of two different phonemes depending on who the listener thinks they are listening to (Johnson et al., 1999). When linguistic and speaker cues conflict, listeners are slower at categorizing speech sounds into phonemes (Apfelbaum et al., 2014), indicating an integration of talker and linguistic information in speech processing. These findings have implications on speech perception: familiarity with a talker's voice results in improved word and sentence intelligibility, improved recognition memory, and decreased processing time (Clarke & Garrett, 2004; Theodore et al., 2015). These benefits are heightened in noisy situations (Nygaard & Pisoni, 1998), highlighting, once again, that in real-life contexts, knowing who is speaking leads to efficient communication.

Voice recognition does not come easily to all individuals. One group that may have difficulty with aspects of talker processing is cochlear implant (CI) users. A CI is a sensory aid for individuals with severe-to-profound sensorineural hearing loss. CI devices function by gathering information about the fluctuation of sound energy over time within frequency bandwidths, converting them into electrical impulse patterns, and directing them to specific electrodes located along the cochlea. While the CI does not restore acoustic hearing, it provides recipients with hearing sensitivity within the speech range. Many individuals with CIs receive gains in decoding linguistic information, including in speech perception and word recognition (e.g., Blamey et al., 2012; N. R. Peterson et al., 2010). With experience, CI users become adept at discriminating voices from environmental sounds (Massida et al., 2011). Some children with CIs even perform comparably to their NH peers when categorizing human vocalizations from environmental sounds (Berland et al., 2019). Nonetheless, it has been argued that there are limitations on the use of cochlear implants for talker processing.

Why might this be? The electric speech signal transmitted via modern CIs are degraded, especially when compared to the acoustic speech signal (see Baskent et al., 2016 for review). Different CI devices employ different processing strategies to transform the speech signal, but a common transformation strategy relies on filtering the acoustic signal into bands of frequencies, resulting in blurred frequency variations within bandpass channels. Moreover, there are physiological limitations of electrical stimulation of the auditory nerve. The auditory nerve responds differently to electric stimulation than acoustic stimulation. Further, the spatial overlap of the broad stimulation from individual electrodes leads to blurred spatial activation patterns. Thus, it is not yet possible to achieve very fine-tuned simulation points with electrodes.

This degradation of the speech signal can make a host of speech and sound perception tasks more difficult for CI users (Ciocca et al., 2002; Hopkins & Moore, 2009). Specifically, to distinguish voices, NH listeners rely on a combination of different acoustic cues, but fundamental frequency (F0) and cues to vocal tract length (VTL) appear to be the most helpful for talker perception (Skuk & Schweinberger, 2014). Both of these cues rely on the harmonic and formant structure of speech, but how they are encoded by the CI device and how they are perceived by CI users is quite limited. For example, pitch can be coded through simulation rate, temporal pattern of stimulation, or place of stimulation – but, even among these coding strategies, the percept of pitch is consistently reported as being weak (Moore & Carlyon, 2005). Likewise, the blurred spectral resolution appears to negatively affect CI users' ability to perceive VTL cues (Gaudrain & Başkent, 2015). Taken together, device and physiological constraints appear to not be conducive for talker perception.

Beyond the ability to encode acoustic cues, access to a language's sound structure assists talker recognition by allowing listeners to distinguish between variability in speech sounds and variability in different talker's voices (see Creel & Bregman, 2011 for review). This in turn helps individuals track and adapt to the idiosyncrasies of a person's voice. Indeed, prior work has shown a gradient influence of phonological processing on talker recognition tasks. Listeners have heightened talker recognition skills for talkers who speak their native language, compared to talkers with a different accent (Vanags et al., 2005) or talkers who speak the listener's second language (Bregman & Creel, 2014). Further, individuals who perform worse on phonological processing tasks, such as individuals with dyslexia, tended to have difficulty with talker recognition (Kadam et al., 2016; Perrachione et al., 2011).

Based on these findings, the predictions for the performance of CI users on talker processing tasks are clear. The degradation of the speech signal should make talker processing a challenging task for CI users, especially compared to listeners with typical hearing. Here, we review the literature to systematically compile the evidence. How well do CI users perform on tasks of talker discrimination, voice perception, and talker identification, compared to their normal-hearing peers? What factors (e.g., participant, hearing, device) affect their performance on these tasks? Is their performance on these tasks related to their performance on other linguistic tasks?

We chose to focus on studies with participants who are CI users, instead of focusing on experiments that used CI-simulated stimuli to investigate talker processing (i.e., through degrading stimuli with noise vocoding). Because NH participants are easier to recruit, presenting degraded stimuli can serve as a first step in uncovering interesting avenues of investigation for improving CI outcomes (Krull et al., 2012). These studies provide a useful foundation to further examine perceptual skills of CI users. However, these studies have two limitations. First, the vocoded manipulations that researchers use in these experiments are based on the manipulations that CI devices do in their processors, but what CI users actually experience were reported to differ than how the simulations sound. Thus, while vocoded stimuli are a good approximate, they are not always analogous. Second, CI users and NH listeners rely on different cues for speech or talker perception since their prior experience with the speech signal differs (e.g., Fuller et al., 2014). On that note, it is important to acknowledge that CI users are a heterogeneous group. For example, CI users can differ by chronological age, device characteristics, onset and duration of deafness, length of CI use, and communication mode - each of which have been shown to contribute to a CI user's level of success in speech perception (Belzner & Seal, 2009; Roberts et al., 2013). It remains a question of whether the same effects hold for talker processing.

Of particular interest, CI users can also have different configurations of devices, which will provide them with different levels of access to sounds. Individuals with unimodal CIs (whether unilaterally, or bilaterally) typically have no residual acoustic hearing in either ear, and thus must rely exclusively on the electric stimulation from their implant. Conversely, individuals with bimodal or hybrid CIs may have residual acoustic hearing. In these contexts, a hearing aid or the hybrid CI could amplify low-frequency sounds. If talker processing relies extensively on the fundamental frequency, then we would expect improved voice recognition for both bimodal and hybrid CI users. But, if talker processing also relies on higher frequency spectral information, then bimodal or hybrid CI configurations might not result in performance levels equal to those of NH participants.

In the present review, we will summarize the current knowledge on talker processing in individuals with cochlear implants, with a focus on talker discrimination and identification. Reflecting the heterogeneity of CI users in the general population, these studies vary in the participants that they have recruited (i.e., in age, device configuration, and hearing experience). Further, they vary in the methodologies that they use to assess talker processing. This has produced a wealth of evidence in various ways, which we now aim to analyze to extract overall coherent findings. First, we scanned the literature, and we tracked the following characteristics in order to identify any research gaps in the literature:

(1) ***Participant characteristics:*** As previously mentioned, there is large heterogeneity in the hearing experiences of CI users. Here, we summarized the characteristics of CI users who

have taken part in studies on talker processing in order to identify which groups the current results are relevant for, and to identify potential gaps in the literature.

(2) ***Task characteristics:*** Talker processing can be assessed in various ways, and prior work has suggested that different paradigms can lead to different conclusions (see Perrachione, 2017). Here, we reviewed the different tasks that are being used with CI users, with a focus on the task paradigms and the types of stimuli being used. We use the term talker processing to refer to the broad topic of this review. When referring to results that apply to a task, we specify talker discrimination, talker identification or voice perception. The breakdown of these categories is described in the Results section.

Further, we reviewed the literature to examine the following research questions:

(1) ***Comparison with NH listeners:*** NH listeners are often used as control groups in experiments assessing CI users' performance. Thus, we reviewed the literature to examine the performance of CI users, as a group, in talker processing tasks, compared to NH listeners.

**(2)** ***Comparison between device configurations:*** Here, we reviewed the studies that examine whether different hearing configurations impact the manner with which CI users process talker information. If residual acoustic hearing can help with voice perception, then we would expect improved talker processing for bimodal CI users, compared to unimodal users. If access to these low-frequency sounds drive voice perception, we might even expect bimodal CI users to perform as well as NH listeners. In contrast, if talker processing also relies on high-frequency spectral information, then bimodal or hybrid CI configurations would not result in performance levels equal to those of NH participants.

**(3)** ***Role of different acoustic cues:*** NH listeners take advantage of various cues, such as F0 and acoustic correlates of VTL (e.g., distribution of formant peaks), to encode talker identity (Smith & Patterson, 2005). Given that CI users do not have the same access to these cues, an open question is whether CI users make use of different cues for identifying voices. Here, we report the findings from studies that investigated how various acoustic cues influence talker information processing in CI users.

**(4)** ***Relationship to participant characteristics:*** Tracking the trajectory of talker processing for different CI users is important for developing expectations about speech processing performance. However, this task is complicated by the wide heterogeneity in demographic and hearing experiences of CI users. Indeed, individuals with different hearing experience, such age at implantation, can have different hearing outcomes (e.g., Manrique et al., 2004). For example, in an investigation of melodic contour identification, Tao and colleagues (2015) found that post-lingual CI users outperformed pre-lingual users. Here, we explored how developmental age and hearing experience (including age of onset of deafness, age at implantation, and duration of CI use) might map onto talker processing abilities.

***(5) Relationship to linguistic tasks:*** Prior work has shown that linguistic and indexical processing is intertwined in speech processing (e.g., Creel & Bregman, 2011). While some individuals might broadly perform well across a variety of speech tasks, it is equally likely that certain strengths and weaknesses will arise within the broad domain of speech processing. Thus, it is of interest to examine how performance in talker processing tasks might relate to their performance in other linguistic tasks, such as consonant/vowel perception or word/sentence recognition.

## 2. Methods

### *2.1. Search strategy*

Figure 1 summarizes the procedural outcomes of this systematic review. The authors ran several searches using the software Publish Or Perish (version 7.15.2643.7260; Harzing, 2007) on Google Scholar. The authors used the search terms "cochlear implant" and every combination of "voice", "talker", and "gender", with either "identification", "generalization", or "discrimination". The searches took place between February 12, 2019 and March 3, 2019. This resulted in 1499 potentially relevant articles. After removing duplicates, 1124 articles remained for further review. The same searches were run again between May 13, 2020 and May 22, 2020 to find any new

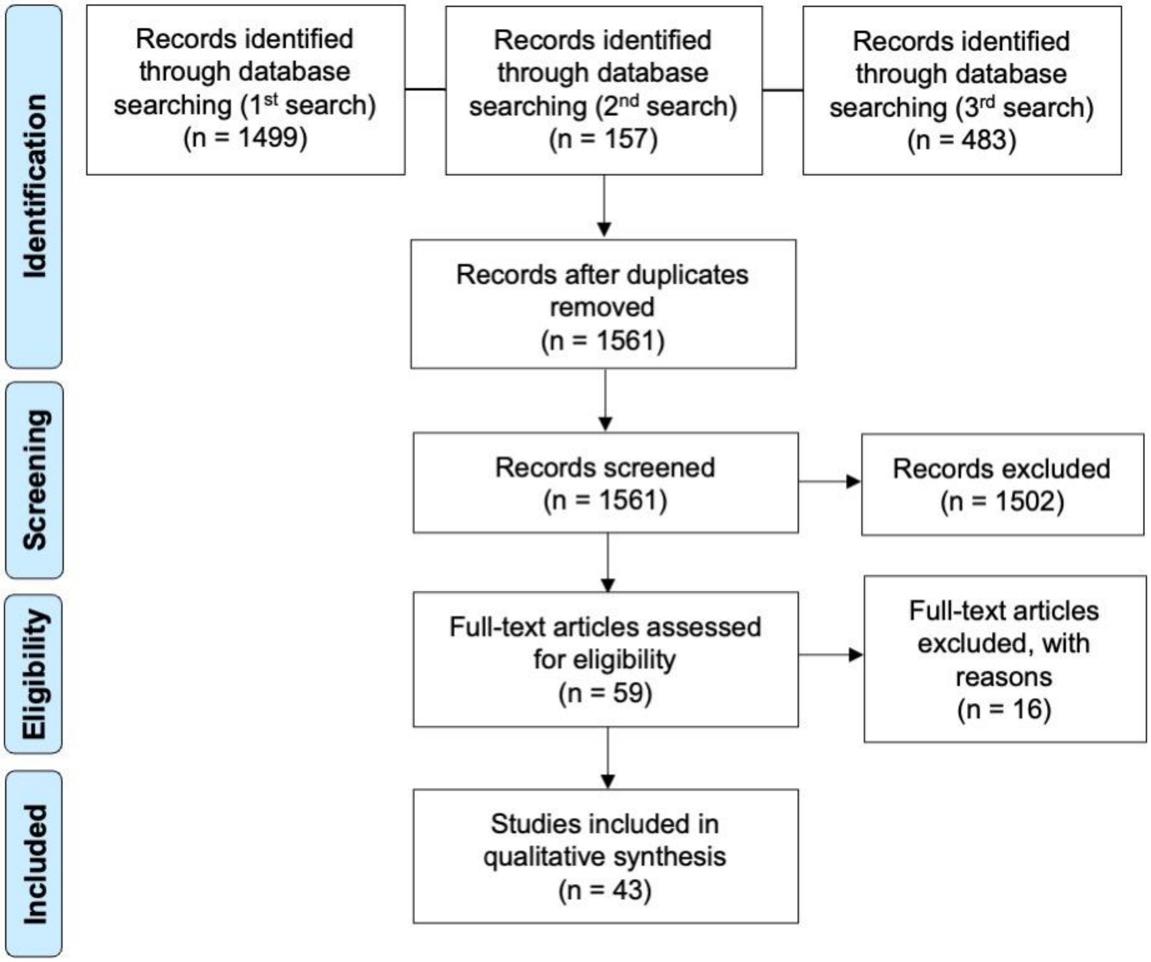

**Figure 1.** PRISMA chart of the study selection process.

research published in the last year. These searches returned 157 results, which resulted in an additional 115 articles to review after duplicates were removed. This led to a total of 1239 articles.

The following inclusion criteria was used to identify relevant articles. We included (1) only primary reports with an experimental study (excluding reports in specialized format, such as theses, dissertations, secondary reports, and conference abstracts); (2) only reports written in English; (3) studies that have at least one group of participants that use cochlear implants; (4) a measure of talker or gender perception, either via an identification or discrimination task. No restrictions were imposed on patient characteristics.

### *2.2. Coding the studies*

In the first triage, the authors scanned the titles and abstracts of 1239 records. Each author was responsible for reviewing half of the records. Articles were coded as either “Include”, “Exclude”, or “Maybe” following our criteria laid out in the previous section. Because the initial triage was based on the title and abstract alone, some articles were ambiguous in whether they met our inclusion criteria (i.e., did not explicitly state if their participants were CI users or did not clearly fit our task criteria). These were coded as “Maybe” and were reviewed by both authors on the basis of title and abstract, and a mutual decision was made regarding their inclusion status. Following review of the abstracts, 48 articles were included for further review.

Each author was responsible for independently reviewing the full-text of half of the remaining articles. Upon more careful review, nine additional articles were excluded after reading the full text for not meeting the inclusion criteria. This left 39 studies in our full review. Participant characteristics (sample size, age, onset of deafness, duration of deafness, age of implantation, length of use, hearing configuration) and study characteristics (goals, outcome measures, stimuli, experimental task, major findings) were summarized for each study. Note that some of these descriptive data were not directly available from the text of manuscripts. In cases where participant-level data was available in tables, we calculated their averages (see footnote of Tables 1, 2 and 3). Two research assistants independently checked the tables for content accuracy. Any clarification or uncertainty was discussed between authors and research assistants until a mutual decision was reached. Studies are reported in three separate tables based on their experimental task: Table 1 summarizes the participant and study characteristics of the studies investigating talker discrimination, Table 2 summarizes the voice perception studies, and Table 3 summarizes the talker identification studies.

### *2.3. Additional search*

Per the suggestion of a reviewer, we conducted a third search with the additional search terms of “voice perception”, “talker perception”, and “gender perception”, along with “cochlear implant”. This search returned 483 articles, and after removing duplicates, resulted in 332 unique articles. Of these, 12 articles were identified as being relevant after reviewing abstracts. After a full-text review, only 4 of these articles were included in the systematic review. We thus include 43 articles in this review.

### *2.3. Quality assessment*

To assess the quality level of the selected studies, we used a methodological quality appraisal tool based on the Grades of Recommendation, Assessment, Development, and Evaluation Working Group (GRADE) approach (Higgins & Green, 2006). Specifically, we adapted an appraisal tool, developed by Downs & Black (1998), for this particular systematic review. The following questions were asked: **Q1)** Was the objective of the study clearly defined? **Q2)** Was the participant inclusion criteria clearly described? **Q3)** Are the main study findings, as pertains to talker discrimination or identification, clearly stated? **Q4)** Are the main outcome measures, as pertains to talker discrimination or identification, clearly stated? **Q5)** Were the investigators blinded to the participant characteristics to reduce bias? **Q6)** Is there a clarification for the appropriateness of the sample size studied? Note that this scale is not necessarily assessing the quality of the evidence, rather our ability to interpret findings based on what was presented in the manuscripts.

Each author assessed ~75% of the selected studies (30 papers), such that approximately 50% of the papers were assessed by both authors (21 papers). Each study was given one point for each "Yes" to the questions above, for a total of 6 points. Studies that received 0-2 points were categorized as "Weak"; studies that received 3-4 points were categorized as "moderate", and studies that received 5-6 points were categorized as "strong". Thus, a "weak" manuscript is one that did not include details that would allow for clear interpretation, while a "strong" study is one that included details that allows for clear interpretation. In total, there were 3 "weak" studies, 24 "moderate" studies, and 16 "strong" studies. Inter-coder ratings were consistent, as none of the rating categories among the papers reviewed by both authors differed. The ratings are included in each study's entry in their respective tables (Table 1, 2, or 3).

## 3. Results

### *3.1. Participant characteristics*

Tables 1, 2, and 3 display the various characteristics of participants in studies covered by this systematic review. Records included a total of 1165 participants with CIs and 381 participants with NH. Certain sets of studies acknowledge being subgroups of one another, resulting in overlapping samples, which means that there were fewer than 1546 unique participants overall. The reviewed studies tended to have small sample sizes, with a median sample size of 15. Group sample sizes for each study are reported in Tables 1, 2, and 3.

In general, there were two main groups of CI users being tested: those focused on adults (i.e., mean age over 18 years of age; n = 30), and those focused on children (i.e., mean age under 18 years of age; n = 13). As is typical in the CI literature, studies with adult participants were mostly focused on middle to late adulthood with postlingual deafness (see Figure 2). For studies with adults that report these characteristics, the mean onset of deafness was 38.5 years of age (SD = 11.87) and the mean age of implantation was 48.59 years of age (SD = 13.94). There was only one study that specifically focused on adult CI users with a prelingual onset of deafness (Zaltz et al., 2018). Participants in this study had an onset of deafness before 1.5 years, but a range of implantation ages from 2.25 - 33.3 years old. Thus, this study is the only study to include early deafened, late implanted individuals. One study had a mix of prelingual and postlingual adult CI users, with one child CI user (Skuk et al., 2020), although these factors were not a focus of investigation. Studies with child participants focused on older children or adolescents with prelingual deafness. The study with the youngest participants was conducted by van Heugten et al. (2014), who recruited children ages 4- to 7-years old. For child studies that report these characteristics, the mean onset of deafness was 0.56 years of age (SD = 0.30) and the mean age of implantation was 3.75 years of age (SD = 2.37). As will be discussed later, these descriptive analyses reveal several gaps of research, including a lack of research on several age ranges that is typical in the CI literature (i.e., children, adolescents, and early adulthood).

Among unimodal CI users, seven studies included only participants with unilateral CIs and one study included only participants with bilateral CIs. Seven studies included a mixed group of unilateral and bilateral CIs. There were three studies that only recruited bimodal CI users, and nine studies recruited a mix of bimodal and unimodal CI users. The remaining twelve studies did not report information about the CI configuration of their participants.

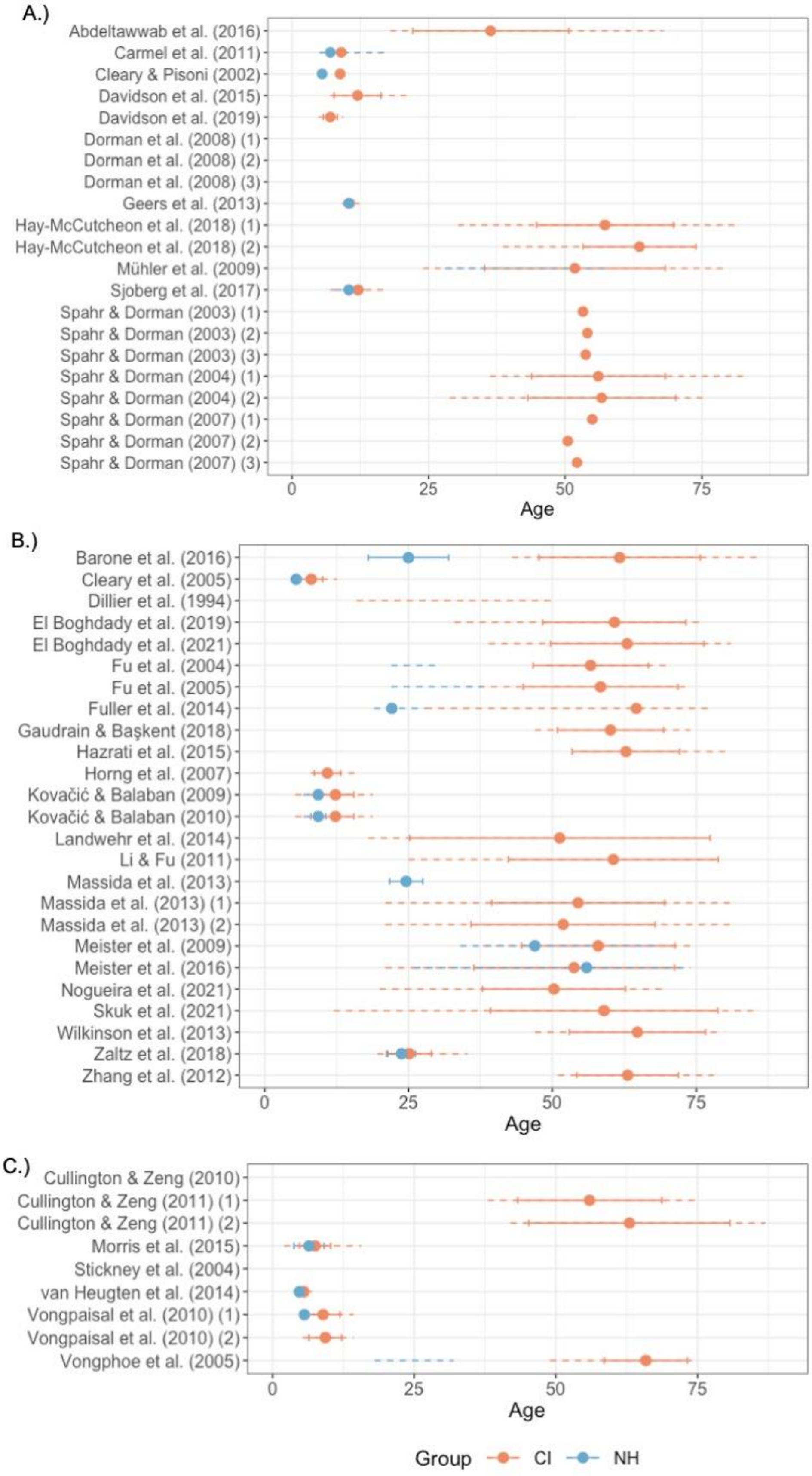

**Figure 2.** Descriptive statistics of the age of participants in papers with A) talker discrimination tasks, B) voice perception tasks, and C) talker identification tasks. Filled circle dots indicate the mean age, solid horizontal lines indicate the standard deviation of age, and dotted horizontal lines indicate the age range. Orange dots and lines represent data for CI participants, while blue dots and lines represent data for NH participants. Numbers in parentheses following the citation indicate different CI groups within the same study.

### *3.2. Task characteristics*

Researchers have noted that different talker processing tasks vary in their sensitivity of detecting differences between groups (if any exists, e.g., Levi, 2019; Perrachione, 2017). In this review, we separated studies into one of three categories: talker discrimination, voice perception, and talker identification. Note, however, that Carmel et al. (Carmel et al., 2011) included both a talker discrimination and talker identification task, and Abdeltawwab et al. (Abdeltawwab et al., 2016) included both a talker discrimination and voice perception task. These categories were informed by the goals of the relevant experimental task. For example, if a study uses a discrimination task with a continua of stimuli along an acoustic dimension, and the main finding is in regards to sensitivity to acoustic cues, this falls under voice perception. In this section, we briefly introduce the tasks that were used in the reported studies, as well as the types of stimuli used, to provide context when discussing findings in later sections.

In *talker discrimination* tasks (n = 15; see Table 1), participants were assessed on their ability to tell different voices apart. This includes both discrimination between and within genders. One study assessed this indirectly by using a parental survey (Carmel et al., 2011), wherein parents were asked if their child had difficulty discriminating between voices. Other laboratory studies generally used a same/different task, in which participants heard two different stimuli and had to respond whether these two tokens were spoken by the same person or by different people. A major feature of this task is that listeners do not need to encode or recall information beyond what is presented in a single trial.

In *voice perception* tasks (n = 21; Table 2), participants were assessed on their ability to perceive differences in vocal cues. This was perhaps the broadest grouping, but these studies all required listeners to perceive differences between talkers or between continua of manipulated talkers. For the most part, these studies are concerned with how CI users perceive acoustic features (F0 and VTL) that cue gender differences, although one study examines cues to vocal age (Skuk et al., 2020). In many cases, these tasks used a two-alternative forced choice paradigm in which participants heard a stimulus and had to respond whether the token was more likely to come from a female or male talker. In one study, participants were asked to judge the “femaleness” or “maleness” of the talker’s voice using a rating scale (Meister et al., 2016). Several studies (El Boghdady et al., 2019, 2021; Gaudrain & Başkent, 2018; Nogueira et al., 2021) used an adaptive three-alternative forced choice task to measure the threshold at which participants could perceive the difference between acoustic cues to vocal gender (i.e., just noticeable differences).

In *talker identification* tasks (n = 7; see Table 3), participants were assessed on their ability to identify voices based on previous experience. These include both recognition of familiar or trained voices and identification of voices into discrete categories. Two studies used parental surveys to assess participants’ talker recognition skills (Carmel et al., 2011; Morris et al., 2015). Other lab experiments used anywhere from a 2- to 10-alternative forced choice task, typically involving a training and a test phase. During training, participants learned different face-voice or name-voice pairings, in which they received feedback regarding their accuracy. During the test phase, participants were asked to identify the person who was speaking. In this cognitively more demanding task, listeners have to encode the relevant features of voices from unfamiliar individuals, hold these in memory, and recall them at a later test phase. Indeed, young children tend to succeed in talker or gender discrimination tasks, but less so in unfamiliar talker identification tasks (Fecher et al., 2019). Given the differences across these tasks, we are careful to label the task when discussing findings below.

Across the different laboratory tasks, different types of stimuli are also used. The tokens themselves were real words (e.g., Consonant-Vowel-Consonant words; n = 13), or short sentences (n = 19). Other studies used isolated vowels (n = 2), non-words (n = 6) or a combination of words and sentences (n = 1). The vast majority of these tokens were produced naturally and used as stimuli in their natural form (n = 29), while other studies used synthetic or synthesized speech (n = 12). For example, Barone et al. (2016) recorded two voices (one female and one male) producing a word, and using a specialized speech manipulation software, they morphed these two voices to a continuum of 11 voices, ranging from 100% female to 100% male and 9 gender-interpolated voices in between.

In the vast majority of cases, CI users' accuracy in talker discrimination, voice perception, and talker identification were above chance levels. There are certainly some studies showing individual differences in performance (to be discussed in more detail later). For example, Kovačić et al. (2010) found that a subset of their CI participants had above-chance performance at identifying the gender of voices, while another subset had below-chance performance. Nonetheless, the fact that most CI users are reaching above-chance levels in these variety of tasks indicates that some talker information is encoded through the CI device.

Some studies set out to examine whether certain forms of speech would affect talker processing. These studies found that CI users were better at picking up on indexical information when presented with neutral speech compared to whispered (Hazrati et al., 2015), speech presented through broadband (compared to presented through the telephone; Horng et al., 2007), and with original stimuli (compared to speech transformed with VTL processing algoritms, Wilkinson et al., 2013). These studies highlight how different real-life situations might impact talker processing for CI users.

### *3.3. Comparison to individuals with typical hearing*

A large proportion of studies (n = 22) included comparisons between CI users and NH listeners. When differences were detected between groups, results showed that CI users had more difficulty in the tasks than NH participants. For example, compared to NH participants, CI users had lower accuracy performance in discriminating between unfamiliar talkers' voices (Carmel et al., 2011; Geers et al., 2013), in identifying the gender of voices (Kovačić & Balaban, 2009; Meister et al., 2009) and in learning to identify voices (van Heugten et al., 2014; Vongpaisal et al., 2010). Differences in talker discrimination appear to be more pronounced when tasks used speakers of the same gender (Muhler et al., 2009), or when using varied stimuli within trials (Cleary et al., 2005; Cleary & Pisoni, 2002). CI users also had more difficulty identifying the gender of voices, compared to NH participants (Massida et al., 2013). In some cases, the best performing CI users overlapped in performance with NH listeners (Cleary et al., 2005; Fu et al., 2005).

Other studies found significant differences between groups in the manner by which listeners encoded vocal properties (Barone et al., 2016; Kovačić & Balaban, 2009; Meister et al., 2016; Zaltz et al., 2018). Another study examined whether top-down information, such as visual information, would influence listeners' gender rating of voices (Barone et al., 2016). Findings from this study show that, compared to NH participants, CI users were more affected by visual information when distinguishing between male and female voices.

Some studies compared CI users' performance to NH listeners' performance with CI simulations (vocoded speech), rather than with unmanipulated speech (Sjoberg et al., 2017; Stickney et al., 2004). In these cases, CI users performed comparably to NH listeners exposed to stimuli that were filtered to a certain amount of channels or frequency bands. For example, Vongphoe et al. (2005) found that CI users performed comparably to NH listeners exposed to stimuli filtered to 1-band through amplitude modulation, while Fu et al. (2005) presented NH listeners with 4- or 8-bands, and van Heugten et al. (2014) with 24-bands.

### *3.4. Comparison between device configurations*

An important question is whether the talker processing abilities of CI users are affected by the configuration of their CI devices. For instance, some studies were interested in the potential advantage of acoustic hearing through hearing aids (HA) to provide indexical cues for talker processing. Cullington and Zeng did not find any significant differences between bimodal users and bilateral CI users on either the identification of specific talkers or the identification of talkers into categories (Cullington & Zeng, 2011, 2010). On the other hand, Hay-McCutcheon et al. (Hay-McCutcheon et al., 2018) found some benefit of acoustic hearing in a talker discrimination task, as bimodal CI users outperformed unilateral CI users. Davidson et al. (2019) found that bimodal children with longer HA use and better Pure Tone Averages (PTA) had a higher suprasegmental speech perception score (a composite score that included performance on talker discrimination), suggesting that prolonged acoustic experience is beneficial as long as hearing loss is not too severe.

Given that between-group comparisons may introduce some confounds, a better test of the hypothesis that residual acoustic hearing may improve voice perception is if bimodal CI users show improved voice perception abilities when listening to combined electric-acoustic stimuli versus just receiving electric or acoustic input. Surprisingly, these studies did not find any significant difference in talker discrimination abilities across the three conditions (Davidson et al., 2015; Dorman et al., 2008). Zhang et al. (2012) administered auditory training to bimodal CI participants, and also found no significant difference in magnitude of improvement in their gender identification performance between electric-alone trials and electric-acoustic stimulation trials. In a talker discrimination task, Abdeltawwab et al. (Abdeltawwab et al., 2016) found an advantage for bimodal stimulation compared to acoustic alone (i.e., HA only), but not compared to electrical alone (i.e., CI only). Taken together, there appears to be minimal evidence that residual acoustic hearing aids in voice perception.

Several studies compared the effects of different CI devices or processing strategies on voice perception. Some studies found that talker processing abilities were not affected by electrode configuration or speech coding strategies (Landwehr et al., 2014). Nevertheless, others have found some key differences in performance based on which processing strategy CI participants were using (i.e., processors from Cochlear Limited, Advanced Bionics, and Med El; Spahr et al., 2007; Spahr & Dorman, 2004, 2003). For instance, Fuller et al. (Fuller et al., 2014) found that users with devices that had a higher stimulation rate were more likely to categorize stimuli as female across a continuum of voices. Geers et al. (2013) found better performance for children who used the most recent CI processor in their study, and Dillier et al. (Dillier et al., 1994) found that the strategy that preserved speech quality features and had a high continuous stimulation rate (High Spectral Transmission) resulted in the best performance in a voice perception task.

### *3.5. Role of different acoustic cues*

NH listeners take advantage of various cues, such as fundamental frequency (F0; related to the pitch of the voice) and the distribution of formant peaks (related to VTL and thus to the height of the talker) to encode talker identity (e.g., Smith & Patterson, 2005). Several studies have investigated how various acoustic cues might influence talker information processing in CI users, including contrasting F0 and VTL, and the spectral and temporal cues that comprise F0. In general, CI users have poorer access to fine acoustic detail, as evidenced by CI users having larger *Just*

*Noticeable Differences* for both F0 and VTL cues compared to NH listeners (Gaudrain & Başkent, 2018).

Given the nature of CI processing, it follows that CI users would adapt to rely more on F0 than on the complex VTL cues. Indeed, findings show that CI users rely more heavily on F0 than VTL for voice gender perception compared to NH listeners, who strongly weigh both cues (Fuller et al., 2014; Meister et al., 2009; Skuk et al., 2020). For instance, CI users are capable of categorizing stimuli from an artificial F0 continuum similarly to NH listeners (Meister et al., 2009) and the discriminability of voices by CI users is correlated to the difference in F0 of the voices, with larger differences in F0 being easier for CI users to discriminate (Muhler et al., 2009). Fu et al. (2005) found that CI users can take advantage of temporal periodicity cues to discriminate gender when voices have distinct F0. However, if the F0 of the voices overlap, CI users were not able to effectively discriminate between genders. It has been suggested that voices that differ by more than an octave, like would be found between a typical male and typical female voice, should be discriminable by CI users, but differences of less than an octave results in ambiguity (see Moore & Carlyon, 2005 for a review of pitch processing in CI users).

Accurate VTL perception requires perception of formant peaks which are more obscured by the limited spectral resolution of a CI. There is some evidence that CI users can use combined F0-VTL cues in sentence contexts (Meister et al., 2016), but it is clear from the limitations of CI processing that it is easier for CI users to access F0 information. For prelingually-deafened adult CI users, age at implantation was related to use of VTL cues, but not F0 (Zaltz et al., 2018). Those who were implanted earlier than age 4 had improved VTL discrimination, suggesting early exposure is important for taking advantage of VTL cues.

A recent study (Skuk et al., 2020) found evidence that CI users do use timbre cues to perceive age differences in voices. This study used a combination of aperiodicity, spectrum level, and formant frequencies to represent timbre and found that high-performing CI users were able to use timbre and F0 to judge age in a similar patter to NH controls (although the controls still outperformed the CI users). The poorer performing CI users, however, showed no evidence of being able to use timbre and consistently judged gender differences as age differences (female voices were categorized younger, while male voices were categorized older).

In summary, CI users adapt their use of acoustic cues from the use of spectral (VTL) information to the greater reliance on F0 to compensate for the degradation of their input compared to NH listeners. Some CI users still take advantage of VTL cues, but this is predicted by early auditory experience. For the most part, CI users seem to rely on F0 as a more reliable cue to distinguishing voices.

### *3.6. Relationship to hearing experience*

Individual differences in hearing history might also affect task performance. These variables include age at onset of deafness, age of implantation, and duration of CI use – all of which have been found to affect anatomical and neurophysiological properties of the auditory system. As noted above, the majority of studies carried out were with either postlingually-deafened adults or prelingually-deafened children, so disentangling the effects of onset of deafness from the age of participants is difficult.  Some studies used a mixed group with participants whose age at onset of deafness ranged from early childhood through adulthood (Fu et al., 2004, 2005; Vongphoe & Zeng, 2005), but they did not investigate the role of age at onset of deafness on the performance of participants.

For individuals who were deaf at a young age, much of the discussion on the effects of hearing experience on voice perception surrounds critical periods. There are two possible predictions. First, if early access to acoustic speech allows for the development of the auditory cortex, then we might expect that children with later onset of deafness might have improved voice perception. However, Kovačić and Balaban (2010) found no substantive effect of age at onset of deafness in the gender identification performance of children ages 5 to 18, with the caveat that cochlear implantation happens early enough (i.e., hearing loss is not left untreated).

Another hypothesis is that children who receive their CIs earlier may have improved voice perception, especially if there is a sensitive period for voice perception. Indeed, children who were implanted at a younger age have better talker discrimination (Geers et al., 2013) and children with a shorter duration of deafness before implantation showed better gender identification (Kovačić & Balaban, 2010). These studies have vastly different ranges of implantation age (1 - 3.2 years versus 2.1 - 15.3 years, respectively) which could explain the different findings with respect to age of implantation and duration of deafness. A benefit of early implantation (before age 4) was also seen in adult prelingually-deafened CI users' discrimination of voice cues (Zaltz et al., 2018). One study only found weak effects, albeit in the expected direction (Cleary et al., 2005). Others did not find any effect of age of implantation in their respective tasks (Cleary & Pisoni, 2002; Morris et al., 2015).

Beyond critical periods, some studies suggest that hearing experience in general contributes to talker processing abilities of CI users. As mentioned previously, Davidson et al. (2019) found that longer hearing aid use benefits suprasegmental perception (by their definition, a composite score that includes performance on talker discrimination) in children with CIs as long as their hearing loss is not too profound, providing evidence for the important role of early auditory input for talker discrimination abilities. Massida et al. (2013) also found that the voice perception skills of CI users can improve over time. Specifically, CI users had poor performance when they were tested on the first month of CI use, but improved over time. Only one study specifically looked at the effects of hearing experience on voice perception in older adult CI users, and they found no relation between factors such as age, duration of deafness or duration of CI experience and participants' performance on a gender identification task (Barone et al., 2016).

### *3.7. Relationship to linguistic tasks*

Here, we investigate how performance in talker processing tasks might relate to their performance in other linguistic and cognitive tasks, such as word recognition, vowel/consonant perception, pitch/prosody perception, speech-on-speech masking, and other cognitive tasks.

#### *3.7.1 Word recognition*

An important question is whether talker processing abilities are related to overall word recognition abilities in CI users. Some studies do not find a statistical relationship between performance in a speech recognition task and a talker processing task. For example, Cullington and Zeng (2011) found no relationship between talker identification performance and scores on a standardized word recognition in noise task (Hearing-in-Noise Test). Similarly, Massida et al. (2013) found no relationship between gender categorization and word recognition performance. These studies suggest that improvements in gender identification or talker identification are independent of improvements in word recognition.

That said, there is some evidence that word recognition abilities are related to talker and voice cue discrimination. Sjoberg et al. (2017) found that talker discrimination was related to speech recognition abilities in adult CI users, and Cleary and Pisoni (2002) found the same relationship in children. Zaltz et al. (2018) found a moderate relationship between the ability to discriminate VTL cues and speech recognition in adult prelingually-deaf CI users. Davidson et al. (2019) found that a composite suprasegmental perception score (that incorporated talker discrimination ability) accounted for variation in both receptive vocabulary and language scores in children. Specifically, the better a child could perceive suprasegmental information, the stronger their receptive language and vocabulary skills would be (measured by the CELF-4 and PPVT-4 respectively). More work is required to investigate the cause of this relationship, but these studies suggest that talker discrimination and word recognition skills are intertwined in speech processing.

### *3.7.2 Vowel and consonant perception*

Vongphoe and Zeng (2005) did not find evidence for a relationship between talker and vowel perception. VTL is closely tied to formant peaks, as are vowels, so it is perhaps surprising that no relationship was found. However, VTL seems to be particularly difficult for CI users (at least compared to F0) so perhaps that CI users seem to move away from relying on VTL is the cause of the dissociation between these skills.

However, Li and Fu (2011) finds that, under certain conditions, vowel and consonant perception is correlated to voice gender discrimination. Particularly, recognizing speech sounds in noise was positively correlated to gender discrimination when talkers only differed by a small average F0 (10 Hz). This finding suggests that some CI users can pick up on subtle acoustic differences in the signal, and take advantage of this in different ways (i.e., recognizing a voice or recognizing phonemes).

### *3.7.3 Speech masking*

Several studies investigated the relationship between speech-on-speech understanding and sensitivity to vocal cues or talker identification (Cullington & Zeng, 2011; El Boghdady et al., 2019, 2021; Nogueira et al., 2021). In a speech-on-speech task, listeners must track a target voice that is masked by a competing voice (or voices). If the target and masker voices are similar, the task becomes more difficult. El Boghdady et al. (2019) found that, on average, CI users with better access to vocal cues (i.e., smaller JNDs to both F0 and VTL) have better performance recognizing single-talker masked speech. Cullington & Zeng (2011), on the other hand, found no correlation between comprehension of single-talker masked sentences and talker identification ability. These conflicting findings suggest that better perception of acoustic cues to voice can benefit speech comprehension in challenging conditions, but that more difficult talker-based tasks do not necessarily relate to one another.

#### *3.7.4 Other abilities*

Several stand-alone studies have investigated various other abilities that might relate to talker processing, including musical experience, cognitive abilities, and affective prosody discrimination. Zaltz et al. (2018) found no relationship between scores on a variety of cognitive tasks (auditory working memory, visual attention, task switching, and nonverbal intelligence) and ability to perceive vocal cues, suggesting that these cognitive functions are not related to discriminability of acoustic features in experienced adult CI users. Cullington & Zeng (2011) found no correlation between talker identification and a standardized assessment of music ability (the Montreal Battery of Evaluation for Amusia). However, prior music experience and musical pitch perception was positively correlated to talker discrimination in children with CIs (Sjoberg et al., 2017). With increasing music experience, CI users performed better at talker discrimination and pitch perception tasks, suggesting that music training can improve pitch perception in CI users which may then enhance processing of talker information (Sjoberg et al., 2017).

Cullington and Zeng (2011) also investigated the relationship between affective prosody discrimination (that is, discrimination of prosody for emotions) and talker identification. There was a strong positive correlation between these two abilities, suggesting that there is at least some relationship between talker identification abilities and learned prosodic processing abilities, like emotion. More work is required to pin down the nature of these relationships, both to further investigate the role of broader cognitive abilities on talker processing and to confirm whether musical training can improve talker perception.

## 4. Discussion

As a group, CI users performed above chance in talker discrimination and identification tasks, suggesting that they are fairly capable processors of talker information. However, there are clear differences in performance when compared to individuals with normal hearing. Compared to NH participants, CI users are less accurate in these talker processing tasks, are less sensitive to the acoustic cues to voice (F0 and VTL), and they use different cues to encode and retrieve indexical properties of voices. These differences highlight the challenges and adaptability of CI users in processing others' voices via electric stimulation.

One clear finding from this review is that there are wide individual differences in performance among CI users. A few studies report that a subset of CI users reach the same level of performance in talker processing tasks as NH participants. Some factors appear to correlate with heightened performance in talker processing tasks. For example, there appears to be some benefit of maintaining residual acoustic hearing for processing talker information. Particularly, bimodal CI users (who have a contralateral HA) tended to perform better at talker discrimination tasks. Additionally, several hearing factors - such as age of onset of deafness, age of implantation, and duration of hearing aid use - is related to talker processing abilities. These findings suggest that early intervention and maintaining acoustic input may be beneficial for talker processing.

Across the studies evaluated here, discrimination abilities relate more clearly to other factors. Particularly, talker discrimination abilities were tied to pitch perception, music experience, and speech recognition abilities in both adults and children (Cleary & Pisoni, 2002; Sjoberg et al., 2017), but the same relationships were not apparent for talker identification abilities. The disparity in findings suggest that identification and discrimination might draw on different perceptual skills. Indeed, it has been suggested that performance in discrimination paradigms rely more on low-level processing of acoustics (see Perrachione, 2017). If such is the case, it may be that this underlying low-level mechanism ties CI users' performance in these different tasks. On the other hand, performance in talker identification tasks has been described as being a better representation of the psychological processes that contribute to voice recognition. It is therefore not surprising that the one significant relationship related to performance in talker identification was with the ability to discriminate affective prosody (Cullington & Zeng, 2011). Further work on this issue in both CI and NH populations would help us better understand the mechanisms that contribute to ecological voice recognition behaviours.

### *4.1. Limitations*

There are several limitations in summarizing this body of literature that makes it difficult to come to firm conclusions. First, many studies reported here fell prey to small samples sizes, which make generalizing their findings difficult. Indeed, the median number of participants for both CI and NH participants in this review was 15. This is an issue that plagues much of clinical language research, as recruiting patients to participate in research can be challenging. In some cases, a small sample size is not a concern if the effect size is large enough for the study to remain well-powered. With a population as variable in outcomes as CI users, larger samples will become necessary when attempting to understand this variability, especially when it comes to investigating interactions between talker processing and broader language or cognitive functioning. In order to predict outcomes for CI users, individual differences approaches require larger samples. The goal of this review was to amalgamate evidence across studies in an attempt to paint a coherent picture of talker processing in CI users.

Second, comparisons between CI and NH participants are often confounded with age. As is typical within the CI literature, the CI participants recruited in the studies listed here tended to be older and included wider age ranges than groups of NH participants (see Figure 2). When direct comparisons are involved, the ages of CI adult participants (M = 56.1 years) were often older than the NH adult participants (M = 33.1 years). Direct comparisons for child participants were more equivalent, the mean ages of child CI participants was 9.5 years compared to 7.4 years for NH child participants. The imbalance in adult comparisons is, in large part, due to the difficulties of recruiting CI participants, as well as in recruiting age-matched participants with normal hearing. Findings from studies with imbalanced ages may be difficult to interpret as there is some evidence that NH older adults perform poorer in talker identification tasks compared to NH younger adults (Best et al., 2018). In other words, it is unclear whether differences in performance between groups are due to age or hearing configuration. When possible, future studies should be more careful about selecting the age ranges of participants.

A related recommendation for researchers is to provide more detailed age and other demographic characteristics of their participants. Four of the studies in this review did not explicitly indicate any age information. Some studies indicate wide age ranges without indicating the mean age, while others indicate the mean age without providing the standard deviation or age range. In addition, several studies provide detailed age information about their participants with

CIs, but lack the same amount of detail for their NH participants. This lack of information makes it difficult to identify the extent to which findings can generalize to other populations.

Lastly, as is apparent from the summary tables in this paper, there is a wide variety of tasks used to assess talker processing. The variability in tasks and outcome measures prohibits the use of a meta-analysis to summarize this work. Nevertheless, as indicated above, these task differences also provide answers to slightly different questions. Discrimination tasks address the differences in talkers that CI users are able to perceive, while identification tasks assess CI users' ability to explicitly label talkers. The difference in perspective between these tasks provides complementary information about talker processing. On one hand, discrimination tasks shed light on the acoustic information that CI users can perceive, while identification tasks address the information that users actually take advantage of to successfully label talkers.

Similarly, this review is limited in the scope of included tasks. We did not include studies whose primary goal was investigating other aspects of talker processing, including speech masking and talker familiarity (although some talker identification studies with training phases may touch on familiarity). Speech masking represents an increased challenge to talker processing which incorporates issues from speech in noise processing, speech streaming, and discriminating competing auditory sources. An extended review of CI performance on these tasks would provide additional framing for how CI users process talker information. Indeed, to successfully recognize a masked talker, one must be able to attend to the target speech stream.

### *4.2. Gaps in research*

This systematic review highlights several gaps in research on indexical processing in CI users. First, while most age ranges are well represented in the literature, there remain gaps of knowledge through the age ranges. There are very few studies examining the talker learning skills of young children, and there is a lack of longitudinal studies within the literature. The studies described in this review provide a solid foundation for understanding the talker learning skills of children with CIs, but they also motivate examining the developmental trajectory of talker learning skills at younger ages. Indeed, prior research has shown that NH infants can recognize highly familiar voices (DeCasper & Fifer, 1980), and that infants can encode and retain unfamiliar voices by 8-months of age (Orena & Werker, 2020). Nonetheless, as with speech perception, the development of voice perception is a gradual process that continues to mature through childhood and adolescence (Creel & Jimenez, 2012; Nagels et al., 2020, 2021; Rigler et al., 2015). Identifying the time points in which young children with CIs become sensitive to indexical information could highlight the plasticity available to child CI users and would be beneficial for clinicians and caregivers of young CI patients.

In a similar vein, there is also a lack of research directly comparing prelingually- and postlingually-implanted CI users. This comparison could provide insight into the different strategies that CI users employ based on their language and acoustic hearing experience. For instance, postlingually-deafened CI users have mental representations of pitch from when they were able to hear acoustic sounds; thus, they may be able to use their memory of acoustic sounds to process pitch through an implant. Prelingual users, on the other hand, have little to no acoustic language experience to draw from, and learn strategies that are informed entirely by input from their implant. Nonetheless, there are a host of confounding variables that could arise when comparing pre- and postlingual CI users. Age, for instance, is likely to be unbalanced across the two groups, as prelingual users will likely be younger than postlingual users.

Towards the same goal of improving outcomes for users, further investigation into the individual differences that predict processing of indexical information could shed light on factors that could be targeted for improvement or enhanced through training. For example, there are mixed results as to the impact of improving talker processing for speech outcomes for CI users, so an experimental intervention study examining this specifically could clarify the issue. There is a wealth of evidence that NH listeners use linguistic information for talker processing (for review, Creel & Bregman, 2011), but that same work is lacking with CI users. The work that does investigate the relationship between linguistic and indexical processing in CI users is correlational in nature and offers inconclusive findings. An experimental study manipulating talker identification in CI users' native compared to non-native language could begin to fill this gap.

We have also highlighted the importance of residual acoustic hearing, but additional demographic and cognitive factors could prove important. Only one study has investigated the role of executive function on voice perception abilities (Zaltz et al., 2018), leaving room for further investigation. Additionally, an examination of the psychophysical skills that relate to talker perception could address whether improvements need to be made to the quality input to improve talker processing. For child CI users, one might also look at how the variety of input received (e.g., how many speakers children interact with) impacts individuals' ability to process indexical information.

### *4.3. Recommendations*

Several studies offer suggestions for improvements to CI processing that could directly benefit talker processing. Broadly, these recommendations include hopes for better algorithms and coding schemes or better optimization strategies for coding algorithms for both children and adults (Geers et al., 2013; Kovačić & Balaban, 2009). For instance, the inclusion of temporal fine structure in CI processors has the potential to improve performance of CI users not just in talker processing, but across a variety of auditory tasks, including speech recognition in noise and perception of music and lexical tone (Stickney et al., 2004). Additionally, VTL perception has been singled out as a limitation imposed by CI technology (Zaltz et al., 2018). Better coding strategies or fitting algorithms that address the shortcomings of VTL discrimination (be it the inability to discriminate VTL or to use it as a cue for talker-size discrimination) could improve talker gender identification for CI users (Fuller et al., 2014). Gaudrain and Başkent (2018) suggests using VTL discrimination as a potential clinical tool for improving CI fit. It could be used as a measure of spectral resolution that is more related to speech perception than spectral ripple tasks. Developing a clinical tool with stronger ties to speech perception could have large implications for CI outcomes related to indexical processing.

Aside from improvements to the CI processor itself, there is also room for improvement in rehabilitation strategies. There is an accumulation of evidence, also supported in this review, that electric-acoustic stimulation (EAS) improves performance for CI users. However, not all CI users have access to residual acoustic hearing. There is some work suggesting that altering the pitch to an audible range for CI users might be beneficial for individuals with limited residual acoustic hearing (Brown et al., 2016), although this work has so far been done with simulations and has yet to be tested with actual CI users. Even further, Huang and colleagues (2017) provide evidence that converting the fundamental frequency of a voice into tactile vibrations (electro-tactile stimulation) improves speech reception thresholds in CI users at a similar magnitude to the benefit seen with EAS. If encoding of talker information cannot be improved at the level of the CI device, then perhaps turning to a different modality (i.e., tactile) to improve new CI users' adaptation to electric input is a worthwhile avenue of investigation.

### *4.5 Conclusion*

The goal of this review was to provide a comprehensive overview of talker processing in CI users. We found that CI users are able to perceive indexical information, but generally have more challenges than their NH peers. Several factors were found to relate to perception of talker information, including residual acoustic hearing and task demands. Future studies could build on this existing work by investigating training paradigms that might improve long-term talker processing abilities. Improved talker processing abilities may then impact other speech outcomes. In the long term, further improvement to devices themselves is crucial to improving talker processing outcomes for CI users.

**Acknowledgements**

S. Colby is supported by NIH grant P50 000242 awarded to B. Gantz and B. McMurray, and A. Orena is supported by a Fonds de Recherche Québec Nature et Technologies (FRQNT) Postdoctoral Fellowship. We thank K. Chen, C. Shum, M. Huffman, and C. Morales for their assistance in coding the studies. We thank B. McMurray for his comments on an earlier draft of this paper.

## References

Abdeltawwab, M. M., Khater, A., & El-Anwar, M. W. (2016). Contralateral bimodal stimulation: A way to enhance speech performance in Arabic-speaking cochlear implant patients. *ORL*, *78*(3), 126–135. https://doi.org/10.1159/000381024

Apfelbaum, K. S., Bullock-Rest, N., Rhone, A. E., Jongman, A., & McMurray, B. (2014). Contingent categorisation in speech perception. *Language, Cognition and Neuroscience*, *29*(9), 1070–1082. https://doi.org/10.1080/01690965.2013.824995

Barone, P., Chambaudie, L., Strelnikov, K., Fraysse, B., Marx, M., Belin, P., & Deguine, O. (2016). Crossmodal interactions during non-linguistic auditory processing in cochlear-implanted deaf patients. *Cortex*, *83*, 259–270. https://doi.org/10.1016/j.cortex.2016.08.005

Baskent, D., Gaudrain, E., Tamati, T. N., & Wagner, A. (2016). Perception and Psychoacoustics of Speech in Cochlear Implant Users. In A. T. Cacace, E. de Kleine, A. G. Holt, & P. van Dijk (Eds.), *Scientific Foundations of Audiology. Perspectives from Physics, Biology, Modelling, and Medicine* (pp. 285–319). Plural Publishing Inc. https://books.google.de/books?hl=de&lr=&id=EtAyDAAAQBAJ&oi=fnd&pg=PA285&dq=Scientific+Foundations+of+Audiology&ots=cfEfTicv7h&sig=1cTQmXsc_FR7oNQiwWYgpklOkN0

Belzner, K. A., & Seal, B. C. (2009). Children with cochlear implants: A review of demographics and communication outcomes. *American Annals of the Deaf*, *154*(3), 311–334.

Berland, A., Collett, E., Gaillard, P., Guidetti, M., Strelnikov, K., Cochard, N., Barone, P., & Deguine, O. (2019). Categorization of everyday sounds by cochlear implanted children. *Scientific Reports*, *9*(1), 1–16. https://doi.org/10.1038/s41598-019-39991-9

Best, V., Ahlstrom, J. B., Mason, C. R., Roverud, E., Perrachione, T. K., Kidd, G., & Dubno, J. R. (2018). Talker identification: Effects of masking, hearing loss, and age. *The Journal of the Acoustical Society of America*, *143*(2), 1085–1092. https://doi.org/10.1121/1.5024333

Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., Dillier, N., Dowell, R., Fraysse, B., Gallégo, S., Govaerts, P. J., Green, K., Huber, A. M., Kleine-Punte, A., Maat, B., Marx, M., Mawman, D., Mosnier, I., O'Connor, A. F., … Lazard, D. S. (2012). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients. *Audiology and Neurotology*, *18*(1), 36–47. https://doi.org/10.1159/000343189

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

Bregman, M. R., & Creel, S. C. (2014). Gradient language dominance affects talker learning. *Cognition*, *130*(1), 85–95. https://doi.org/10.1016/j.cognition.2013.09.010

Brown, C. A., Helms Tillery, K., Apoux, F., Doyle, N. M., & Bacon, S. P. (2016). Shifting Fundamental Frequency in Simulated Electric-Acoustic Listening. *Ear & Hearing*, *37*(1), e18–e25. https://doi.org/10.1097/AUD.0000000000000227

Carmel, E., Kronenberg, J., Wolf, M., & Migirov, L. (2011). Telephone use among cochlear implanted children. *Acta Oto-Laryngologica*, *131*(2), 156–160. https://doi.org/10.3109/00016489.2010.517784

Ciocca, V., Francis, A. L., Aisha, R., & Wong, L. (2002). The perception of Cantonese lexical tones by early-deafened cochlear implantees. *The Journal of the Acoustical Society of America*, *111*(5), 2250. https://doi.org/10.1121/1.1471897

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The*

*Journal of the Acoustical Society of America*, *116*(6), 3647–3658. https://doi.org/10.1121/1.1815131

Cleary, M., & Pisoni, D. B. (2002). Talker discrimination by prelingually deaf children with cochlear implants: Preliminary results. *Annals of Otology, Rhinology and Laryngology*, *111*(5 II), 113–118. http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L34522316%5Cnhttp://sfx.library.uu.nl/utrecht?sid=EMBASE&issn=00034894&id=doi:&atitle=Talker+discrimination+by+prelingually+deaf+children+with+cochlear+implants%3A+Preliminary+result

Cleary, M., Pisoni, D. B., & Kirk, K. I. (2005). Influence of Voice Similarity on Talker Discrimination in Children With Normal Hearing and Children With Cochlear Implants. *Journal of Speech, Language, and Hearing Research*, *48*(1), 204–223. https://doi.org/10.1044/1092-4388(2005/015)

Creel, S. C., & Bregman, M. R. (2011). How Talker Identity Relates to Language Processing. *Linguistics and Language Compass*, *5*(5), 190–204. https://doi.org/10.1111/j.1749-818X.2011.00276.x

Creel, S. C., & Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*, *113*(4), 487–509. https://doi.org/10.1016/j.jecp.2012.07.007

Cullington, H. E., & Zeng, F.-G. (2011). Comparison of Bimodal and Bilateral Cochlear Implant Users on Speech Recognition with Competing Talker, Music Perception, Affective Prosody Discrimination, and Talker Identification. *Ear and Hearing*, *32*(1). https://doi.org/10.1097/AUD.0b013e3181edfbd2

Cullington, H. E., & Zeng, F. G. (2010). Comparison of bimodal and bilateral cochlear implant users. *Cochlear Implants International*, *11 Suppl 1*, 67–74. https://doi.org/10.1179/146701010X12671177440262

Davidson, L. S., Firszt, J. B., Brenner, C., & Cadieux, J. H. (2015). Evaluation of Hearing Aid Frequency Response Fittings in Pediatric and Young Adult Bimodal Recipients. *Journal of the American Academy of Audiology*, *26*(4), 393–407. https://doi.org/10.3766/jaaa.26.4.7

Davidson, L. S., Geers, A. E., Uchanski, R. M., & Firszt, J. B. (2019). Effects of Early Acoustic Hearing on Speech Perception and Language for Pediatric Cochlear Implant Recipients. *Journal of Speech, Language, and Hearing Research*, *62*(9), 3620–3637. https://doi.org/10.1044/2019_JSLHR-H-18-0255

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176.

Dillier, N., Lai, W., & Bögli, H. (1994). A high spectral transmission coding strategy for a multi-electrode cochlear implant. *Advances in Cochlear Implants*, 152–157.

Dorman, M. F., Gifford, R. H., Spahr, A. J., & McKarns, S. A. (2008). The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies. *Audiology and Neurotology*, *13*(2), 105–112. https://doi.org/10.1159/000111782

El Boghdady, N., Gaudrain, E., & Başkent, D. (2019). Does good perception of vocal characteristics relate to better speech-on-speech intelligibility for cochlear implant users? *The Journal of the Acoustical Society of America*, *145*(1), 417–439. https://doi.org/10.1121/1.5087693

El Boghdady, N., Langner, F., Gaudrain, E., Başkent, D., & Nogueira, W. (2021). Effect of Spectral Contrast Enhancement on Speech-on-Speech Intelligibility and Voice Cue

Sensitivity in Cochlear Implant Users. *Ear and Hearing*, 271–289. https://doi.org/10.1097/AUD.0000000000000936

Fecher, N., Paquette-Smith, M., & Johnson, E. K. (2019). Resolving the (Apparent) Talker Recognition Paradox in Developmental Speech Perception. *Infancy*, *24*(4), 570–588. https://doi.org/10.1111/infa.12290

Fu, Q.-J., Chinchilla, S., & Galvin, J. J. (2004). The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *JARO - Journal of the Association for Research in Otolaryngology*, *5*(3), 253–260. https://doi.org/10.1007/s10162-004-4046-1

Fu, Q.-J., Chinchilla, S., Nogaki, G., & Galvin, J. J. (2005). Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. *The Journal of the Acoustical Society of America*, *118*(3), 1711–1718. https://doi.org/10.1121/1.1985024

Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q. J., Free, R. H., & Başkent, D. (2014). Gender Categorization Is Abnormal in Cochlear Implant Users. *JARO - Journal of the Association for Research in Otolaryngology*, *15*(6), 1037–1048. https://doi.org/10.1007/s10162-014-0483-7

Gaudrain, E., & Başkent, D. (2015). Factors limiting vocal-tract length discrimination in cochlear implant simulations. *The Journal of the Acoustical Society of America*, *137*(3), 1298–1308. https://doi.org/10.1121/1.4908235

Gaudrain, E., & Başkent, D. (2018). Discrimination of voice pitch and vocal-tract length in cochlear implant users. *Ear and Hearing*, *39*(2), 226–237. https://doi.org/10.1097/AUD.0000000000000480

Geers, A. E., Davidson, L. S., Uchanski, R. M., & Nicholas, J. G. (2013). Interdependence of

Linguistic and Indexical Speech Perception Skills in School-Age Children With Early Cochlear Implantation. *Ear and Hearing*, *34*(5), 562–574. https://doi.org/10.1097/AUD.0b013e31828d2bd6

Harzing, A. W. (2007). *Publish or Perish* (7.15.2643.7260). https://harzing.com/resources/publish-or-perish

Hay-McCutcheon, M. J., Peterson, N. R., Pisoni, D. B., Iler, K., Yang, X., Parton, J., Kirk, K. I., Yang, X., & Parton, J. (2018). Performance variability on perceptual discrimination tasks in profoundly deaf adults with cochlear implants. *Journal of Communication Disorders*, *72*(January 2017), 122–135. https://doi.org/10.1016/j.jcomdis.2018.01.005

Hazrati, O., Ali, H., Hansen, J. H. L., & Tobey, E. (2015). Evaluation and analysis of whispered speech for cochlear implant users: Gender identification and intelligibility. *The Journal of the Acoustical Society of America*, *138*(1), 74–79. https://doi.org/10.1121/1.4922230

Higgins, J. P. T., & Green, S. (2006). Highly sensitive search strategies for identifying reports of randomized controlled trials in MEDLINE. In *Cochrane Handbook for Systematic Reviews of Interventions 4.2.6* (p. Appendix 5b). www.cochrane.org/resources/handbook/hbook.htm

Hopkins, K., & Moore, B. C. J. (2009). The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *The Journal of the Acoustical Society of America*, *125*(1), 442–446. https://doi.org/10.1121/1.3037233

Horng, M., Chen, H., Hsu, C., & Fu, Q. (2007). Telephone Speech Perception by Mandarin-Speaking Cochlear Implantees. *Ear & Hearing*, *28*, 66S-69S.

Huang, J., Sheffield, B., Lin, P., & Zeng, F. G. (2017). Electro-tactile stimulation enhances cochlear implant speech recognition in noise. *Scientific Reports*, *7*(1), 1–5. https://doi.org/10.1038/s41598-017-02429-1

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27*(4), 359–384. https://doi.org/10.1006/jpho.1999.0100

Kadam, M. A., Orena, A. J., Theodore, R. M., & Polka, L. (2016). Reading ability influences native and non-native voice recognition, even for unimpaired readers. *The Journal of the Acoustical Society of America*, *139*(1), EL6–EL12. https://doi.org/10.1121/1.4937488

Kovačić, D., & Balaban, E. (2009). Voice gender perception by cochlear implantees. *The Journal of the Acoustical Society of America*, *126*(2), 762–775. https://doi.org/10.1121/1.3158855

Kovačić, D., & Balaban, E. (2010). Hearing History Influences Voice Gender Perceptual Performance in Cochlear Implant Users. *Ear & Hearing*, *31*, 806–814.

Krull, V., Luo, X., & Iler Kirk, K. (2012). Talker-identification training using simulations of binaurally combined electric and acoustic hearing: Generalization to speech and emotion recognition. *The Journal of the Acoustical Society of America*, *131*(4), 3069–3078. https://doi.org/10.1121/1.3688533

Landwehr, M., Fürstenberg, D., Walger, M., von Wedel, H., & Meister, H. (2014). Effects of various electrode configurations on music perception, intonation and speaker gender identification. *Cochlear Implants International*, *15*(1), 27–35. https://doi.org/10.1179/1754762813Y.0000000037

Levi, S. V. (2019). Methodological considerations for interpreting the Language Familiarity Effect in talker processing. *Wiley Interdisciplinary Reviews: Cognitive Science*, *10*(2), 1–15. https://doi.org/10.1002/wcs.1483

Li, T., & Fu, Q. J. (2011). Voice gender discrimination provides a measure of more than pitch-related perception in cochlear implant users. *International Journal of Audiology*, *50*(8), 498–502. https://doi.org/10.3109/14992027.2011.576274

Manrique, M., Cervera-Paz, F. J., Huarte, A., & Molina, M. (2004). Advantages of cochlear implantation in prelingual deaf children before 2 years of age when compared with later implantation. *Laryngoscope*, *114*(8 I), 1462–1469. https://doi.org/10.1097/00005537-200408000-00027

Massida, Z., Belin, P., James, C., Rouger, J., Fraysse, B., Barone, P., & Deguine, O. (2011). Voice discrimination in cochlear-implanted deaf subjects. *Hearing Research*, *275*(1–2), 120–129. https://doi.org/10.1016/j.heares.2010.12.010

Massida, Z., Marx, M., Belin, P., James, C., Fraysse, B., Barone, P., & Deguine, O. (2013). Gender Categorization in Cochlear Implant Users. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1389–1401. https://doi.org/10.1044/1092-4388(2013/12-0132)

Meister, H., Fürsen, K., Streicher, B., Lang-Roth, R., & Walger, M. (2016). The Use of Voice Cues for Speaker Gender Recognition in Cochlear Implant Recipients. *Journal of Speech, Language, and Hearing Research*, *59*(3), 546–556. https://doi.org/10.1044/2015_JSLHR-H-15-0128

Meister, H., Landwehr, M., Pyschny, V., Walger, M., & Wedel, H. von. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *International Journal of Audiology*, *48*(1), 38–48. https://doi.org/10.1080/14992020802293539

Moore, B. C. J., & Carlyon, R. P. (2005). Perception of Pitch by People with Cochlear Hearing Loss and by Cochlear Implant Users. In C. J. Plack, R. R. Fay, A. J. Oxenham, & A. N. Popper (Eds.), *Pitch. Springer Handbook of Auditory Research, vol 24* (pp. 234–277).

Springer. https://doi.org/10.1007/0-387-28958-5_7

Morris, D. J., Christiansen, L., Uglebjerg, C., Brännström, K. J., & Falkenberg, E. (2015). Parental comparison of the prosodic and paralinguistic ability of children with cochlear implants and their normal hearing siblings. *Clinical Linguistics & Phonetics*, *29*(11), 840–851. https://doi.org/10.3109/02699206.2015.1055803

Muhler, R., Ziese, M., Rostalski, D., Mühler, R., Ziese, M., Rostalski, D., Muhler, R., Ziese, M., & Rostalski, D. (2009). Development of a Speaker Discrimination Test for Cochlear Implant Users Based on the Oldenburg Logatome Corpus. *ORL*, *71*(1), 14–20. https://doi.org/10.1159/000165170

Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., & Başkent, D. (2020). Development of voice perception is dissociated across gender cues in school-age children. *Scientific Reports*, *10*(1), 1–11. https://doi.org/10.1038/s41598-020-61732-6

Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., & Başkent, D. (2021). School-age children benefit from voice gender cue differences for the perception of speech in competing speech. *The Journal of the Acoustical Society of America*, *149*(5), 3328–3344. https://doi.org/10.1121/10.0004791

Nogueira, W., Boghdady, N. El, Langner, F., Gaudrain, E., & Başkent, D. (2021). Effect of Channel Interaction on Vocal Cue Perception in Cochlear Implant Users. *Trends in Hearing*, *25*. https://doi.org/10.1177/23312165211030166

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics*, *60*(3), 355–376. https://doi.org/10.3758/BF03206860

Orena, A. J., & Werker, J. (2020). *Infants use disambiguation to learn new voices*. 1–26. https://doi.org/10.31234/osf.io/czkhx

Perrachione, T. K. (2017). Speaker recognition across languages. *The Oxford Handbook of Voice Perception*. https://open.bu.edu/handle/2144/23877

Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. E. (2011). Human voice recognition depends on language ability. *Science*, *333*(6042), 595. https://doi.org/10.1126/science.1207327

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, *109*(6), 2988–2998. https://doi.org/10.1121/1.1370525

Peterson, G. E., & Barney, H. L. (1952). Control Methods Used in a Study of the Vowels. *Journal of the Acoustical Society of America*, *24*(2), 175–184. https://doi.org/10.1121/1.1906875

Peterson, N. R., Pisoni, D. B., & Miyamoto, R. T. (2010). Cochlear implants and spoken language processing abilities: Review and assessment of the literature. *Restorative Neurology and Neuroscience*, *28*(2), 237–250. https://doi.org/10.3233/RNN-2010-0535

Rigler, H., Farris-Trimble, A., Greiner, L., Walker, J., Tomblin, J. B., & McMurray, B. (2015). The slow developmental time course of real-time spoken word recognition. *Developmental Psychology*, *51*(12), 1690–1703. https://doi.org/10.1037/dev0000044

Roberts, D. S., Lin, H. W., Herrmann, B. S., & Lee, D. J. (2013). Differential cochlear implant outcomes in older adults. *Laryngoscope*, *123*(8), 1952–1956. https://doi.org/10.1002/lary.23676

Sidtis, D., & Kreiman, J. (2012). In the Beginning Was the Familiar Voice: Personally Familiar Voices in the Evolutionary and Contemporary Biology of Communication. *Integrative Psychological and Behavioral Science*, *46*(2), 146–159. https://doi.org/10.1007/s12124-

011-9177-4

Sjoberg, K. M., Driscoll, V. D., Gfeller, K., Welhaven, A. E., Kirk, K. I., & Prusick, L. (2017). The impact of electric hearing on children's timbre and pitch perception and talker discrimination. *Cochlear Implants International*, *18*(1), 36–48. https://doi.org/10.1080/14670100.2016.1263406

Skuk, V. G., Kirchen, L., Oberhoffner, T., Guntinas-Lichius, O., Dobel, C., & Schweinberger, S. R. (2020). Parameter-Specific Morphing Reveals Contributions of Timbre and Fundamental Frequency Cues to the Perception of Voice Gender and Age in Cochlear Implant Users. *Journal of Speech, Language, and Hearing Research*, *63*(9), 3155–3175. https://doi.org/10.1044/2020_JSLHR-20-00026

Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech, Language, and Hearing Research*, *57*(1), 285–296. https://doi.org/10.1044/1092-4388(2013/12-0314)

Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America*, *118*(5), 3177–3186. https://doi.org/10.1121/1.2047107

Spahr, A. J., & Dorman, M. F. (2004). Performance of Subjects Fit with the Advanced Bionics CII and Nucleus 3G Cochlear Implant Devices. *Archives of Otolaryngology - Head and Neck Surgery*, *130*(5), 624–628. https://doi.org/10.1001/archotol.130.5.624

Spahr, A. J., & Dorman, M. F. (2003). A comparison of performance among patients with the CII Hi-Resolution, 3G and Tempo+ processors. *2003 Conference on Implantable Auditory Protheses*, *January*, 5. http://www.zachary.com/creemer/three_group_comparison1.pdf

Spahr, A. J., Dorman, M. F., & Loiselle, L. H. (2007). Performance of patients using different cochlear implant systems: Effects of input dynamic range. *Ear and Hearing*, *28*(2), 260–275. https://doi.org/10.1097/AUD.0b013e3180312607

Stickney, G. S., Nie, K., Kong, Y. Y., Chen, H., & Zeng, F. G. (2004). Temporal fine structure: The missing component in speech processing algorithms. *International Congress Series*, *1273*(C), 23–26. https://doi.org/10.1016/j.ics.2004.09.017

Tao, D., Deng, R., Jiang, Y., Galvin, J. J., Fu, Q. J., & Chen, B. (2015). Melodic pitch perception and lexical tone perception in mandarin-speaking cochlear implant users. *Ear and Hearing*, *36*(1), 102–110. https://doi.org/10.1097/AUD.0000000000000086

Theodore, R. M., Myers, E. B., & Lomibao, J. A. (2015). Talker-specific influences on phonetic category structure. *The Journal of the Acoustical Society of America*, *138*(2), 1068–1078. https://doi.org/10.1121/1.4927489

van Heugten, M., Volkova, A., Trehub, S. E., & Schellenberg, E. G. (2014). Children's recognition of spectrally degraded cartoon voices. *Ear and Hearing*, *35*(1), 118–125. http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L563083740%5Cnhttp://sfx.library.uu.nl/utrecht?sid=EMBASE&issn=15384667&id=doi:&atitle=Children%27s+recognition+of+spectrally+degraded+cartoon+voices.&stitle=Ear+Hear&title=Ear+and+h

Vanags, T., Carroll, M., & Perfect, T. J. (2005). Verbal overshadowing: A sound theory in voice recognition? *Applied Cognitive Psychology*, *19*(9), 1127–1144. https://doi.org/10.1002/acp.1160

Vongpaisal, T., Trehub, S. E., Schellenberg, E. G., Lieshout, P. Van, & Papsin, B. C. (2010). Children with Cochlear Implants recognize their Mother's voice. *Ear & Hearing*, *31*, 555–

566.

Vongphoe, M., & Zeng, F.-G. (2005). Speaker recognition with temporal cues in acoustic and electric hearing. *The Journal of the Acoustical Society of America*, *118*(2), 1055–1061. https://doi.org/10.1121/1.1944507

Wilkinson, E. P., Abdel-Hamid, O., Galvin, J. J., Jiang, H., & Fu, Q. J. (2013). Voice conversion in cochlear implantation. *Laryngoscope*, *123*(SUPPL. 3). https://doi.org/10.1002/lary.23744

Zaltz, Y., Goldsworthy, R. L., Kishon-Rabin, L., & Eisenberg, L. S. (2018). Voice Discrimination by Adults with Cochlear Implants: the Benefits of Early Implantation for Vocal-Tract Length Perception. *JARO - Journal of the Association for Research in Otolaryngology*, *19*(2), 193–209. https://doi.org/10.1007/s10162-017-0653-5

Zhang, T., Dorman, M. F., Fu, Q. J., & Spahr, A. J. (2012). Auditory Training in Patients With Unilateral Cochlear Implant and Contralateral Acoustic Stimulation. *Ear and Hearing*, *33*(6), e70–e79. https://doi.org/10.1097/AUD.0b013e318259e5dd