# Meta-Analysis on the Identification of Linguistic and Emotional Prosody in Cochlear Implant Users and Vocoder Simulations

Marita K. Everhardt,[1,2,3] Anastasios Sarampalis,[3,4] Matt Coler,[5] Deniz Başkent,[2,3] and Wander Lowie[1,3]

**Objectives:** This study quantitatively assesses how cochlear implants (CIs) and vocoder simulations of CIs influence the identification of linguistic and emotional prosody in nontonal languages. By means of meta-analysis, it was explored how accurately CI users and normal-hearing (NH) listeners of vocoder simulations (henceforth: simulation listeners) identify prosody compared with NH listeners of unprocessed speech (henceforth: NH listeners), whether this effect of electric hearing differs between CI users and simulation listeners, and whether the effect of electric hearing is influenced by the type of prosody that listeners identify or by the availability of specific cues in the speech signal.

**Design:** Records were found by searching the PubMed Central, Web of Science, Scopus, Science Direct, and PsycINFO databases (January 2018) using the search terms "cochlear implant prosody" and "vocoder prosody." Records (published in English) were included that reported results of experimental studies comparing CI users' and/or simulation listeners' identification of linguistic and/or emotional prosody in nontonal languages to that of NH listeners (all ages included). Studies that met the inclusion criteria were subjected to a multilevel random-effects meta-analysis.

**Results:** Sixty-four studies reported in 28 records were included in the meta-analysis. The analysis indicated that CI users and simulation listeners were less accurate in correctly identifying linguistic and emotional prosody compared with NH listeners, that the identification of emotional prosody was more strongly compromised by the electric hearing speech signal than linguistic prosody was, and that the low quality of transmission of fundamental frequency ($f_0$) through the electric hearing speech signal was the main cause of compromised prosody identification in CI users and simulation listeners. Moreover, results indicated that the accuracy with which CI users and simulation listeners identified linguistic and emotional prosody was comparable, suggesting that vocoder simulations with carefully selected parameters can provide a good estimate of how prosody may be identified by CI users.

**Conclusions:** The meta-analysis revealed a robust negative effect of electric hearing, where CIs and vocoder simulations had a similar negative influence on the identification of linguistic and emotional prosody, which seemed mainly due to inadequate transmission of $f_0$ cues through the degraded electric hearing speech signal of CIs and vocoder simulations.

**Key words:** Cochlear implants, Perception, Prosody, Vocoder simulations.

(Ear & Hearing 2020;41;1092–1102)

[1]Center for Language and Cognition Groningen, University of Groningen, The Netherlands; [2]Department of Otorhinolaryngology, University Medical Center Groningen, The Netherlands; [3]Research School of Behavioural and Cognitive Neurosciences, University of Groningen, The Netherlands; [4]Department of Experimental Psychology, University of Groningen, The Netherlands; and [5]Campus Fryslân, University of Groningen, The Netherlands.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site (www.ear-hearing.com). Address for correspondence: Marita K. Everhardt, Center for Language and Cognition Groningen, University of Groningen, Oude Kijk in 't Jatstraat 26, 9712 EK Groningen, The Netherlands. E-mail: m.k.everhardt@rug.nl



## INTRODUCTION

Cochlear implants (CIs) are auditory prostheses that can partially restore hearing in individuals with severe to profound hearing loss by using electrodes to stimulate the auditory nerve directly via electric current. However, the speech signal transmitted through a CI is highly lacking in fine spectrotemporal detail, compromising the identification of speech, especially in complex listening scenarios such as noisy environments or multiple-talker situations (Shannon et al. 2004; Pisoni 2005; Başkent et al. 2016b; Plack 2018). Previous research has shown that compared with normal-hearing (NH) listeners of unprocessed speech, CI users or NH listeners of speech degraded with acoustic simulations of CIs through vocoder simulations are less accurate in correctly identifying segmental elements of speech (e.g., Green et al. 2005; Luo & Fu 2009; Luo et al. 2009), speech in noise (e.g., Meister et al. 2011a; van Zyl & Hanekom 2013), speaker gender (e.g., Fu et al. 2004, 2005; Meister et al. 2009; Fuller et al. 2014b), lexical tone in tonal languages (e.g., Lee et al. 2002; Peng et al. 2017; Wang et al. 2011), and prosodic elements of speech (in nontonal languages) such as linguistic prosody and emotional prosody (e.g., Luo et al. 2007; Torppa et al. 2010; Meister et al. 2011b; Chatterjee et al. 2015). To focus on linguistic and emotional prosody in nontonal languages, relatively few studies have examined the influence of CIs or vocoder simulations on prosody identification and this influence has to date—to our knowledge—not yet been quantified. This study aims to provide a quantitative overview of these studies by quantifying the influence of the degraded electric hearing speech signal of CIs and degraded speech due to vocoder simulations on the identification of linguistic and emotional prosody by means of a meta-analysis. This meta-analysis assesses whether the negative influence of CIs or vocoder simulations on prosody identification found in individual studies is a valid and robust finding across studies. The magnitude of the effect will be determined by taking into account differences in the experimental design and sample sizes (i.e., CI research is typically limited in number of participants) of individual studies, providing a more comprehensive overview of how CIs or





<zdoi; 10.1097/AUD.0000000000000863>

vocoder simulations influence the identification of prosody than would be possible in individual studies.

Prosody forms an important part of spoken language. The focus in this study is on two forms of prosody, namely linguistic prosody and emotional prosody in nontonal languages. Linguistic prosody conveys information on the syntactic and semantic properties of speech and contributes to a listener's ability to identify boundaries between syllables and words, as well as affecting the identification of sentences influenced by for instance stress or sentence type. Emotional prosody conveys information on the emotional state of a speaker and contributes to a listener's ability to identify this emotion (note that this can also be achieved through nonauditory cues such as facial expressions (e.g., Most & Aviner 2009; Chatterjee et al. 2015; Fengler et al. 2017), yet that is beyond the scope of the present study as the focus here is on auditory-only studies). These two forms of prosody utilize a common set of acoustic cues—most prominently fundamental frequency ($f_0$), intensity, and duration, but also voice quality and utterance-wide variation in formant spacing—and listeners are able to draw on these cues when identifying the information conveyed by the different forms of prosody (Gussenhoven 2004; Raithel & Hielscher-Fastabend 2004; Ladd 2008; Ladefoged & Johnson 2010; Belyk & Brown 2014; Goedemans et al. 2018). Deficits in the ability to identify prosody—due to for instance inadequate access to acoustic cues—could have serious consequences for social development and social interactions or for speech categorization and language acquisition (e.g., Raithel & Hielscher-Fastabend 2004; Geers et al. 2013; Belyk & Brown 2014; Chatterjee et al. 2015). Compared with NH listeners of unprocessed speech (henceforth: NH listeners), CI users are at a disadvantage in their ability to identify prosody due to inadequate access to some of these cues (note that this study focuses on $f_0$, intensity, and duration only). That is, where NH listeners are able to use $f_0$ decoded by either temporal or place cues, these cues are degraded in the electric hearing signal of CIs (Shannon et al. 2004; Pisoni 2005; Başkent et al. 2016b; Plack 2018). As a result, CI users show reduced ability in prosody identification when it heavily relies on $f_0$. Similarly, though to a lesser extent, CI users also show reduced ability in prosody identification when it heavily relies on intensity (Shannon 2002; Moore 2003; Drennan & Rubinstein 2008; Meister et al. 2011b). At the same time, CI users show an ability comparable to that of NH listeners when identification heavily relies on duration (Shannon 2002; Drennan & Rubinstein 2008; Meister et al. 2011b).

A recent study by van de Velde et al. (2015) was the first—and to our knowledge only—study that directly compared cue-weightings for the identification of linguistic prosody and for the identification of emotional prosody in electric hearing [(also reported in the dissertation of van de Velde (2017)]. The results of the study showed that listeners apply different cue-weighting strategies when identifying emotional prosody than when identifying linguistic prosody; a benefit of one cue over the other was found for emotional but not for linguistic prosody. These cue-weighting strategies for each form of prosody were found to be the same for both the NH listeners of vocoder simulations (henceforth: simulation listeners) and the NH listeners (i.e., control group) of the study. When identifying emotional prosody, both listener groups mainly relied on $f_0$ differences ($f_0$ mean, standard deviation, and range were larger for happy stimuli than for sad stimuli), thus attaching the heaviest weight to the cue $f_0$. When identifying linguistic prosody (i.e., focus at phrasal level), however, there was no significant difference in how heavily listeners relied on $f_0$ differences compared with how heavily they relied on duration differences (focused words were longer than nonfocused words), thus attaching an equal amount of weight to these cues. Given these differences in cue-weighting strategies between the different forms of prosody, the difference in quality of transmission of the various cues through the electric hearing speech signal may lead to differences in how accurately linguistic prosody compared with emotional prosody is identified by CI users or simulation listeners. Indeed, the study by van de Velde et al. (2015) has shown that the simulation listeners were less accurate in correctly identifying emotional prosody than in correctly identifying linguistic prosody; they attached the heaviest weight to $f_0$ when identifying emotional prosody yet $f_0$ (decoded by either temporal or place cues) is degraded in the electric hearing signal (Shannon 2002; Moore 2003; Drennan & Rubinstein 2008; Meister et al. 2011b). This suggests that when CI users or simulation listeners do not adjust their cue-weighting strategies to accommodate the degradation of $f_0$ in the signal by attaching more weight to cues that are not degraded (e.g., duration; sad stimuli were longer than happy stimuli)—similar trade-off relationships have been observed in previous research (e.g., Peng et al. 2012)—they identify emotional prosody less accurately than linguistic prosody.

Furthermore, accommodation to low-quality transmission of cues may differ between CI users and simulation listeners, as the identification of the information conveyed through the electric hearing speech signal is—in addition to the quality of the signal—also influenced by the perceptual and cognitive mechanisms of the listeners. CI users are, for instance, affected by physiological changes resulting from hearing loss, leading to differences in the functioning of the auditory and language processing systems between CI users and simulation listeners (Pisoni 2005; Başkent et al. 2016a,b; Wagner et al. 2019). That said, CI users have more experience with the electric hearing speech signal and may therefore be more accustomed to accommodating the low quality of transmission of certain cues compared with simulation listeners. Yet, CI-user studies and vocoder-simulation studies have revealed similar results for these groups with respect to the accuracy with which they identify prosody (van de Velde 2017), suggesting that despite the differences in functioning of the auditory and language processing systems between these groups and the differences in experience with the degraded signal, the accuracy of identification may be comparable. Vocoder simulations with carefully selected parameters thus seem to provide a good estimate of how prosody may be identified by CI users, with the caveat that they are not capable of revealing all details of underlying mechanism differences.

This article reports the results of a meta-analysis quantifying the influence of CIs and vocoder simulations on the identification of linguistic and emotional prosody in nontonal languages. The main objective of the analysis was to assess how the accuracy with which CI users and simulation listeners identify the information conveyed by the different forms of prosody compares to the accuracy with which NH listeners identify this information (hereafter: effect of electric hearing). It was predicted that CI users and simulation listeners are less accurate in correctly identifying linguistic and emotional prosody than NH listeners, presumably due to degradations in fine spectrotemporal detail (Shannon 2002; Raithel & Hielscher-Fastabend 2004; Shannon et al. 2004; Pisoni 2005; Belyk & Brown 2014; Başkent et al. 2016b; Plack 2018). Moreover, quantitative evidence was sought in support of the methodological suitability of vocoder simulations in studies on the effect of electric hearing by investigating whether vocoder stimulations provide an adequate estimate of the influence of CIs on prosody identification. It was predicted that the effect of electric hearing does not differ between studies with CI user and studies with simulation listeners, which would suggest that—despite the differences in functioning of the auditory and language processing systems between these listener groups and the differences in experience with the degraded signal (Pisoni 2005; Başkent et al. 2016a,b; Wagner et al. 2019)—accuracy scores of simulation listeners for vocoder simulations with carefully selected parameters that are thought to provide an adequate estimation of CI hearing may still be a good model of accuracy scores of CI users. Additional objectives of the analysis were to assess whether the effect of electric hearing differs between linguistic and emotional prosody or between unmanipulated and manipulated stimuli, focusing on the availability of the acoustic cues $f_0$, intensity, and duration. It was predicted that, if the results found by van de Velde et al. (2015) are robust across prosody identification studies, the identification of emotional prosody is even more strongly compromised for the CI users and simulation listeners compared with NH listeners than the identification of linguistic prosody due to the heavy weight that is attached to $f_0$—the cue that is degraded in the electric hearing signal—during the identification of emotional prosody (Moore 2003; Drennan & Rubinstein 2008; Meister et al. 2011b; van de Velde et al. 2015). Last, it was predicted that when prosody is signaled through $f_0$ differences, CI users or simulation listeners are less accurate in correctly identifying prosody compared with NH listeners, whereas when it is not signaled through $f_0$ differences (e.g., if $f_0$ is normalized by flattening the $f_0$ contour or by using a noninformative $f_0$ contour), there will be no significant differences between these groups. This is due to the fact that $f_0$ is degraded in the electric hearing signal, resulting in reduced ability in prosody identification when it is signaled through $f_0$. Yet, when prosody identification is signaled through cues that are not degraded in the electric hearing signal, identification accuracy will be comparable between CI users or simulations listeners and NH listeners.

## MATERIALS AND METHODS

### Literature Search and Inclusion Criteria

Figure 1 outlines the search and selection strategies used in the meta-analysis. Records were found by searching the electronic databases PubMed Central, Web of Science, Scopus, Science Direct, and PsycINFO (January 2018) using the search terms “cochlear implant prosody” (662 hits across all databases) and “vocoder prosody” (198 hits across all databases). After removal of duplicates and of announcements, conference abstracts, indices, reference lists, and other nonrelevant texts, the records were screened according to the following inclusion criteria, selecting only:

- Studies with CI users or simulation listeners. As the main objective of the meta-analysis was to investigate the effect of CIs and of vocoder simulations on prosody identification and not to investigate the effect of hearing impairment in itself or of hearing aids (HAs), studies with hearing-impaired individuals that do not use a CI were excluded. Studies that grouped CI and HA users were also excluded.
- Studies on speech prosody, excluding studies on lexical tone, segmental elements of speech, and music prosody. Given the fact that the influence of musical aptitude or musical training on prosody perception in CI users or simulation listeners was not an objective of the present meta-analysis, studies that included musicians or musical training in speech prosody experiments were only included in the meta-analysis if there was a control nonmusician group or a group that did not receive musical training.
- Experimental studies, excluding theoretical papers and review papers.
- Studies on group data, excluding case studies.
- Records published in English.
- Studies with NH controls, as the effect of electric hearing under investigation in the present meta-analysis is defined as how the accuracy with which CI users and simulation listeners identify the information conveyed by the different forms of prosody compares to the accuracy with which NH listeners identify this information.
- Studies on prosody identification. Any studies looking only at prosody production were excluded. When records consisted of both perception and production experiments, only the perception experiments were included, provided they were identification tasks.
- Studies without any other confounding factors such as white-noise, as perceptual influences on prosody identification accuracy other than that of CIs and vocoder simulations was not an objective of the present meta-analysis. Studies with manipulated stimuli were only included if the stimuli were manipulated along one or more of the acoustic cues $f_0$, intensity, and duration, in which case the manipulations were included as moderator variables (see Moderator Analysis).
- Studies on linguistic or emotional prosody (or both) in nontonal languages, excluding studies on indexical prosody (e.g., gender or speaker identification), as the main objective of the analysis was to investigate the identification of the information conveyed by forms of prosody that provide the listener with details on the syntactic and semantic properties of speech and on the emotional state of the speaker, and not on the identity of the speaker. Specifically, the aim was to investigate the accuracy with which CI users and simulation listeners identify prosody irrespective of speaker characteristics.

In total, 29 records met these criteria. By carefully reviewing the reference lists of each of these 29 records in search of previously unidentified eligible records (i.e., records that were not found by searching the electronic databases), five additional records were identified which fit the inclusion criteria. A total of 34 eligible records were subsequently assessed for the availability of statistical data required for the calculation of effect sizes (i.e., means and standard deviations of identification accuracy scores and sample sizes of the groups). If the data were not reported in the paper, the authors were contacted for additional information. Only studies with all the statistical data available—either reported in the article or provided by the authors—were included in the meta-analysis (n = 28).

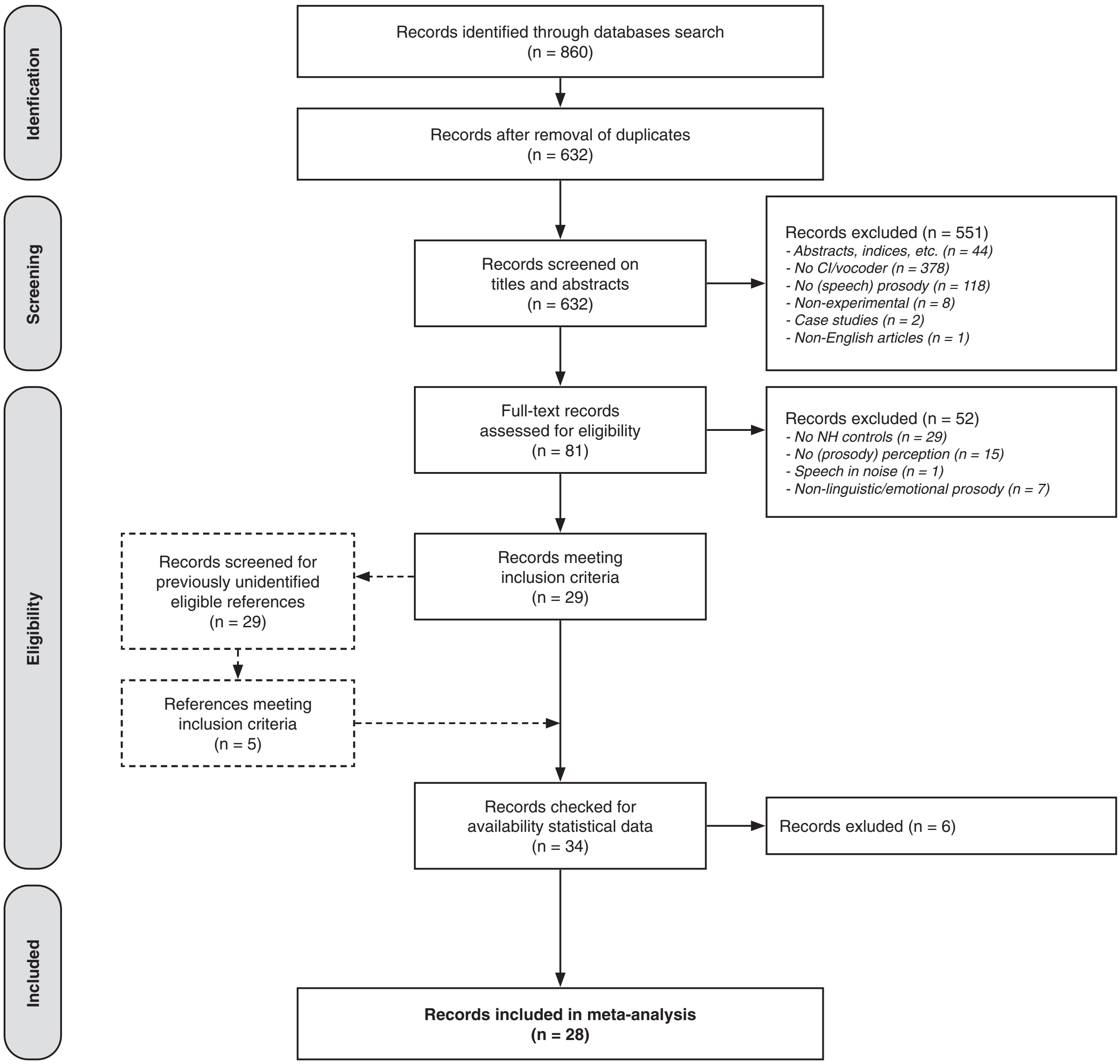


Fig. 1. Flowchart illustrating the search and selection strategies used to select the records to be included in the meta-analysis.

Many of the records selected for inclusion in the meta-analysis consisted of multiple studies, each containing its own condition or comparison. Breaking down each record into its relevant component studies—only including those components that met the inclusion criteria discussed above—resulted in a total of 64 studies available for analysis. These component studies all measured prosody perception by means of a single interval forced choice (1IFC) identification task with the number of alternative forced choices (AFCs) ranging from 2 to 6. Various measures of identification scores were used (e.g., percentage correct, proportion correct, *d*'). Stimuli consisted of single words, phrases, or complete sentences and covered a wide range of nontonal languages. In some studies, the stimuli were manipulated along one or more of the cues $f_0$, intensity, and duration. Details of the experimental design of all records and their relevant component studies can be found in Table, Supplemental Digital Content 1, http://links.lww.com/EANDH/A631.

## Effect Size Calculations

Means and standard deviations of the identification accuracy scores and sample sizes of the groups (i.e., CI users and/or simulation listeners and the NH listeners) were extracted (either from the paper or provided by the authors) for each relevant component study of the records outlined in Table, Supplemental Digital Content 1, http://links.lww.com/EANDH/A631. Effect sizes were calculated as standardized mean differences in identification accuracy scores between groups using the *d* family; as Cohen's *d* gives a biased estimate of the effect size, especially for studies with small sample sizes, effect sizes were calculated as the unbiased corrected effect size Hedges' *g* (Borenstein et al. 2009; Cumming 2012; Lakens 2013). The unbiased correction Hedges' *g* was computed from Cohen's *d* using correction factor *J* and the variance of Hedges' *g* ($V_g$) was computed from the variance of Cohen's *d* ($V_d$) using the squared correction factor *J* ($J^2$; Borenstein et al. 2009).

Effect sizes for the between-subject design studies included in the present meta-analysis (i.e., studies comparing scores of CI users with those of NH listeners) were calculated as Hedges' $g_s$ from Cohen's $d_s$, with the pooled standard deviation as the standardizer of Cohen's $d_s$ (Borenstein et al. 2009; Lakens 2013). The degrees of freedom (*df*) of correction factor *J* for between-subject design studies was $n_1 + n_2 - 2$, where $n_1$ and $n_2$ represent the sample sizes of the two groups (Borenstein et al. 2009). Effect sizes for the within-subject design studies included in the meta-analysis (i.e., studies comparing scores of NH individuals listening to vocoded speech to scores of those same NH individuals listening to unprocessed speech) were calculated as Hedges' $g_{av}$ from Cohen's $d_{av}$, with the average standard deviation of both measures as the standardizer of Cohen's $d_{av}$ (Cumming 2012). This is the optimal effect size calculation for within-subject design data when the correlation (*r*) between the dependent measures is not available (Lakens 2013). The *df* of correction factor *J* for within-subject design studies was $n - 1$, where *n* represents the number of pairs (Borenstein et al. 2009).

### Statistical Analyses

Traditional meta-analytic models assume that there is no dependency between effect sizes and that each study contributes only one effect size to the model (Borenstein et al. 2009; Konstantopoulos 2011; Cheung 2015). Such models can be considered as two-level models with two sources of variance: sampling variance of the observed effect sizes at level 1 and variance between studies at level 2 (Cheung 2014, 2015; Assink & Wibbelink 2016). It is, however, possible for a single study to contribute multiple correlated effect sizes to the model, such as in studies with multiple outcomes or multiple comparison groups (Borenstein et al. 2009; Konstantopoulos 2011; Cheung 2015). In this case, the dependency between the effect sizes can be dealt with by allowing for correlation between effect sizes within a single study (or specified cluster) through the addition of a third level to the meta-analytic model. Such a multilevel meta-analytic model with three sources of variance models sampling variance of the observed effect sizes at level 1, variance between effect sizes from the same study or cluster (i.e., within-study variance) at level 2, and variance between studies or clusters (i.e., between-study variance) at level 3 (Konstantopoulos 2011; Cheung 2014, 2015; Assink & Wibbelink 2016).

The present meta-analysis included effect sizes of 64 component studies extracted from 28 records; some records contributed only one effect size to the model whereas others contributed multiple effect sizes (see Table, Supplemental Digital Content 1, http://links.lww.com/EANDH/A631). For the records contributing multiple effect sizes to the model, the dependency between the effect sizes was either related to the overlap in participants (when the same participants performed different tasks; e.g., Kalathottukaren et al. 2015) or to the overlap in experimental design (when different participant groups performed the exact same task; e.g., Chatterjee et al. 2015). Furthermore, in some cases, there was dependency between effect sizes from different records, as multiple records from the same first author were included in the meta-analysis in which there was participant overlap. This was the case for the records by Meister et al. (2009, 2011b) and for the records by van Zyl and Hanekom (2013) and van Zyl (2014). Records with overlap in participants were pooled into a single cluster. The records by Fuller et al. (2014) and by Gilbers et al. (2015) were therefore also pooled into one cluster, given the large overlap in participants as reported by Gilbers et al. (2015). The records by Peng et al. (2008, 2012), however, were not pooled into one cluster, as there was no overlap in participants, nor in experimental design. This resulted in 28 records being pooled into 25 clusters. To account for the dependency between the effect sizes of studies from the same cluster, the effect sizes were subjected to a multilevel meta-analytic model with the within-study variance between effect sizes from the same cluster modeled at level 2 and the between-study variance between clusters modeled at level 3.

The multilevel random-effects (MLRE) model was modeled with restricted maximum-likelihood estimation using the *rma.mv* function of the *metaphor* package (version 2.1-0; Viechtbauer 2010b) in the R environment. The within-study variance and between-study variance were added as random effects to the model. The use of a random-effects approach is warranted because of the large variability in experimental design between the studies (Thompson & Higgins 2002; Field & Gillett 2010) and the assumption that the selected studies are a random sample of studies (Borenstein et al. 2009; Viechtbauer 2010a). The test statistics were based on the *t*-distribution (Knapp & Hartung 2003; Viechtbauer 2010a; Assink & Wibbelink 2016). The distribution of the variance over the three levels of the multilevel models was determined using the formulas of Cheung (2014) implemented into R syntax by Assink and Wibbelink (2016). The significance of the heterogeneity of within-study variance (at level 2) and between-study variance (at level 3) was determined by one-sided log-likelihood-ratio-tests, comparing the fit of the full three-level model with the fit of a reduced two-level model excluding either the within-study variance or the between-study variance (Assink & Wibbelink 2016).

**Influential Outliers** • The effect sizes of the MLRE model were evaluated for potential influential cases and outliers, as they could affect the validity and robustness of the results (Viechtbauer & Cheung 2010). Influential cases were defined in terms of the diagonal elements of the hat matrix (i.e., hat values). Outliers were defined in terms of standardized residuals. Effect sizes were removed if they were identified as both an influential case and as an outlier (i.e., influential outliers). Influential outliers were identified if the hat values were greater than two times the average hat value and if the standardized residual values exceeded three standard deviations from the mean (Stevens 1984; Viechtbauer & Cheung 2010; Aguinis et al. 2013).

**Missing Data** • The validity of the results of a meta-analysis is highly dependent on the underlying data and is sensitive to missing data. Missing data can, for instance, be due to publication bias, which occurs when the publication of research depends on the statistical significance of the results of the study. That is, studies reporting statistically significant results are more likely to be published than studies reporting nonsignificant results. Missing data can also be due to other forms of bias that influence the search process, such as language bias, availability bias, duplication bias, or citation bias (Hopewell et al. 2005; Borenstein et al. 2009). Bias in meta-analyses is associated with funnel plot asymmetry (Egger et al. 1997). The possibility of missing data in the present meta-analysis due to any form of bias was evaluated with Egger's regression test, which measures funnel plot asymmetry (Egger et al. 1997; Sterne & Egger 2005). Egger's regression test was conducted by adding the standard error (SE) of the effect sizes as a moderator to the MLRE model. A significant estimate of the intercept was considered evidence for funnel plot asymmetry, which would indicate that there was missing data in the meta-analysis and that the results of the MLRE model would be influenced by bias.

**Moderator Analysis** • The MLRE model was extended to a mixed-effects model to investigate whether predetermined variables can be identified as moderators of the effect size estimate. That is, in case of significant within-study variance and between-study variance, the variability in effect sizes cannot solely be attributed to sampling variance. A mixed-effects model can be used to determine whether the variance at the different levels of the MLRE model could be explained by any (or all) of the moderator variables (Assink & Wibbelink 2016). Furthermore, the mixed-effects model was also used to explore the additional objectives identified for the present meta-analysis (see Introduction), namely whether the effect of electric hearing differs between CI users and simulation listeners, between linguistic and emotional prosody, or between unmanipulated and manipulated stimuli, focusing on the availability of the acoustic cues $f_0$, intensity, and duration. The following factors were therefore included as moderator variables in the mixed-effects model: (a) group: CI users versus simulation listeners, (b) prosody: linguistic versus emotional, (c) stimuli: unmanipulated versus manipulated, (d) $f_0$: unmanipulated versus normalized, (e) intensity: unmanipulated versus normalized, and (f) duration: unmanipulated versus normalized.

Note that age is often an inherent factor in CI research, as two main groups of CI users are children who are prelingually deafened and implanted at a young age and older individuals who are postlingually deafened and implanted at a relatively advanced age (Blamey et al. 2013; Tobey et al. 2013; Dunn et al. 2014). Implanted children will develop their auditory perceptual skills via the implant, which will be shaped by degraded sound input, as well as the neural plasticity period during childhood (Sharma et al. 2005). Often, under ideal conditions of no comorbidity and no other health-related complications, these children develop strong linguistic skills, yet, likely their perceptual skills differ from both NH children and postlingual CI adults. For older adults, speech perception may be affected by age-related perceptual and cognitive changes, yet, especially in vocoder studies, often there is a discrepancy in age ranges of participant groups, due to difficulties in recruiting normal-hearing older adults (Bhargava et al. 2016). In this meta-analysis of prosody identification studies, however, age was not included as a moderator variable, for several reasons. First, this was not the focus of the analysis. But also, second, in an earlier version of the mixed-effects model, when the factor age (categorized into children studies and adult studies) was included as a moderator variable, there was no significant influence on the effect size estimate and age as a factor was therefore excluded. Similarly, language was also not included as a moderator variable, as this was also not the focus of the analysis. Although the studies included in this meta-analysis cover a wide range of nontonal languages and the focus of the analysis was the identification of prosody in nontonal languages, the question of how this may differ between different nontonal languages was not an objective identified for the present meta-analysis. Moreover, in an earlier version of the mixed-effects model, when the factor language was included as a moderator variable, there was also no significant influence on the effect size estimate and language as a factor was therefore also excluded.

The multilevel mixed-effects (MLME) model with the factors group, prosody, stimuli, $f_0$, intensity, and duration as moderator variables was modeled with restricted maximum-likelihood estimation using the *rma.mv* function of the *metaphor* package (version 2.1-0; Viechtbauer 2010b) in the R environment. The intercept of the MLME model was set to reflect the effect size of studies in which CI users identify emotional prosody with unmanipulated stimuli. In addition, to determine the significance of the moderating effect of individual moderator variables, separate MLME models each including only one factor were modeled. The influence of moderator variables on the effect size estimate was determined with an omnibus test; the test yields a significant result if at least one regression coefficient deviates from zero. If the coefficient of a variable significantly deviates from zero, that variable has a significant moderating effect on the effect size estimate. The omnibus tests of the MLME models followed the *F*-distribution (Knapp & Hartung 2003; Viechtbauer 2010a; Assink & Wibbelink 2016). The distribution of the variance and the significance of the heterogeneity of within-study variance (at level 2) and between-study variance (at level 3) was determined in the same manner as for the MLRE model.

## RESULTS

Figure 2 illustrates the effect sizes of the influence of CIs and vocoder simulations on the identification of prosody (i.e., effect of electric hearing) for each component study. Negative effect sizes denote a negative effect of electric hearing on prosody identification, meaning that prosody identification with electric hearing is less accurate than prosody identification with normal acoustic hearing. Positive effect sizes denote that prosody identification is more accurate with electric hearing than with acoustic hearing. The MLRE model (k = 64) revealed a combined Hedges' *g* effect size estimate of –1.84 [SE = 0.21, 95% CI (–2.25, –1.43)], a large effect size which reached significance [$t(63) = -8.95$, $p < 0.001$]. This indicates that electric hearing had a negative effect on prosody identification and thus that CI users and simulation listeners identified linguistic and emotional prosody less accurately than NH listeners.

Assessment of the distribution of the total variance over the three levels of the MLRE model showed that 10.4% of the total variance can be attributed to the sampling variance at level 1, that 44.1% of the total variance can be attributed to differences in effect sizes from the same cluster (i.e., within-study variance) at level 2, and that 45.5% of the total variance can be attributed to differences between clusters (i.e., between-study variance) at level 3. The one-sided log-likelihood-ratio tests revealed significant variance at both level 2 ($p < 0.001$) and level 3 ($p < 0.001$), indicating that the variability in effect sizes cannot solely be attributed to sampling variance. The extension of the MLRE model to the MLME was therefore warranted; the MLME model can determine whether the within-study or between-study variance can be explained by moderator variables.

Outlier and influential case analyses did not detect any influential outliers in the dataset. The analysis did identify two component studies (Chatterjee et al. 2015, study no. 2.2; Luo et al. 2007, study no. 10.1) as outliers (i.e., standardized residual values of –3.83 and –3.38, respectively), yet neither of these outliers was identified as influential because the effect

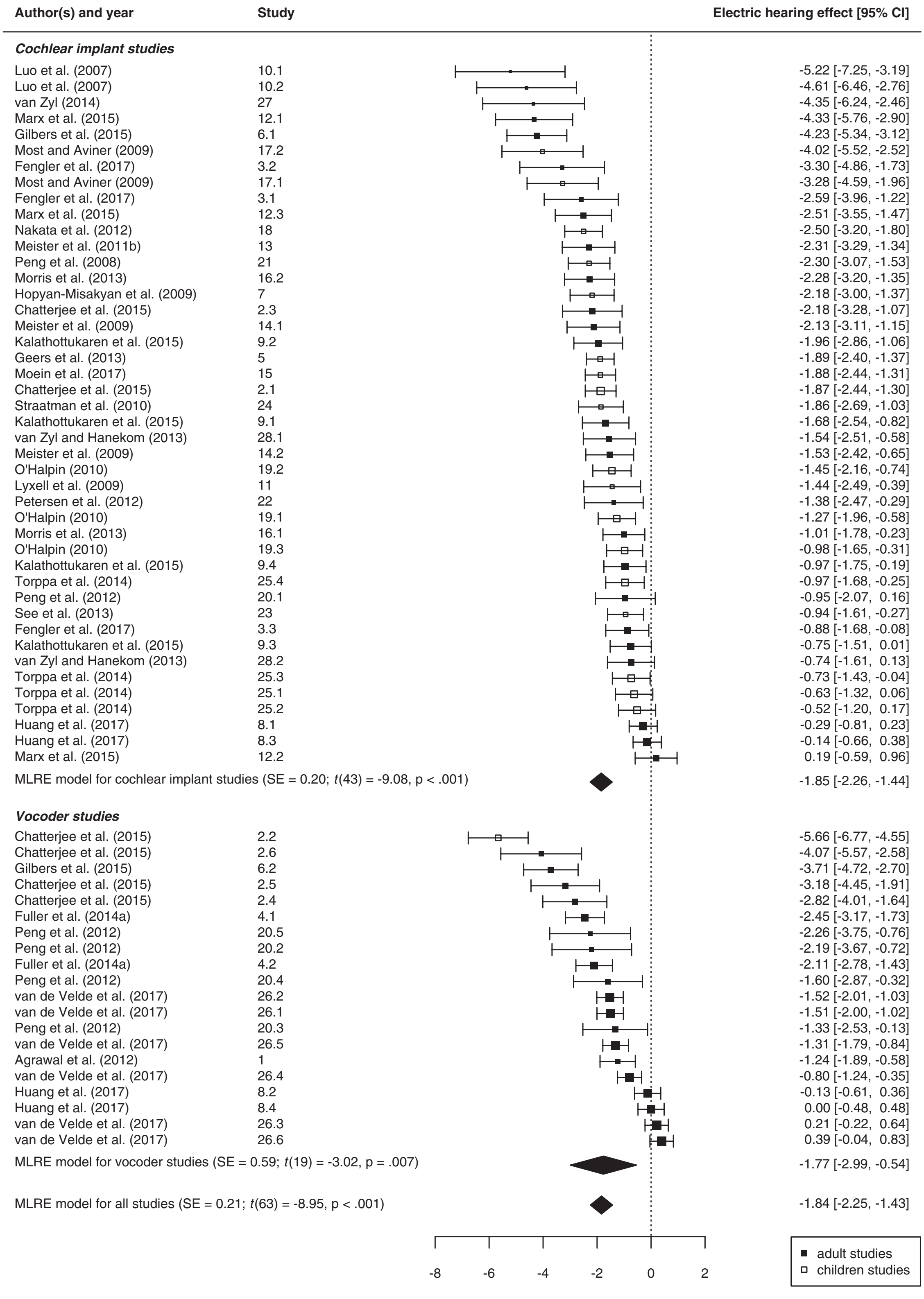


Fig. 2. Effect sizes and 95% CIs of studies included in the meta-analysis. Symbol size reflects sample size. Combined Hedges' $g$ effect size estimates, 95% CI, SE, and test statistics of the MLRE models are shown for cochlear implant studies, for vocoder studies, and for all studies. CI indicates confidence interval; MLRE, multilevel random-effects; SE, standard errors.

size hat values of these studies did not exceed two times the average hat value (i.e., 0.57 and 0.89, respectively). The validity and robustness of the results of the MLRE model were therefore not affected by influential outliers. The Egger's regression test, however, did reveal asymmetry in the data [$t(62) = 2.11$, $p = 0.039$], indicating that there was missing data and that the findings of the MLRE model were influenced by bias.

## Moderator Variables

The omnibus test of the MLME model (k = 64) with the factors group, prosody, stimuli, $f_0$, intensity, and duration as moderator variables revealed that the effect size estimate was moderated by at least one of the included variables [$F(6, 57) = 5.05$, $p < 0.001$]. The intercept (i.e., set to reflect the effect size of studies in which CI users identify emotional prosody with unmanipulated stimuli) significantly deviated from zero [–2.33, SE = 0.27, 95% CI (–2.86, –1.80); $t(57) = -8.77$, $p < 0.001$]. The MLME model revealed two factors that significantly influenced this effect size estimate: studies with linguistic prosody decreased the magnitude of the effect size by 0.76 [SE = 0.28, 95% CI (0.19, 1.32); $t(57) = 2.68$, $p = 0.010$] and studies in which $f_0$ of the stimuli was normalized decreased the magnitude of the effect size by 1.79 [SE = 0.58, 95% CI (0.62, 2.96); $t(57) = 3.07$, $p = 0.003$]. Both linguistic prosody and normalized $f_0$ thus had a robust and unique moderating effect on the effect of electric hearing. The other factors (i.e., group, stimuli, intensity, and duration) had no significant influence on the effect size estimate, as none of the regression coefficients of these factors significantly deviated from zero. This finding was corroborated by the omnibus tests of the separate MLME models for these factors, as none of these yielded a significant result.

Important to note is that the test for residual heterogeneity of the MLME model also reached significance [$Q_E(57) = 287.20$, $p < 0.001$], indicating that there is significant unexplained variance left. Of the total variance of the MLME model, 14.2% can be attributed to sampling variance, 33% to within-study variance, and 52.9% to between-study variance. The one-sided log-likelihood-ratio tests revealed significant within-study variance ($p < 0.001$) and significant between-study variance ($p = 0.001$). The moderator variables included in this MLME model thus did not account for all the variance between effect sizes, suggesting that there are additional moderator variables not accounted for in this MLME model that are likely to be important. This is, however, beyond the scope of this study.

**Effect of Prosody** • The omnibus test of the MLME model (k = 64) with only the factor prosody confirmed that the effect size estimate was moderated by the type of prosody listeners had to identify [$F(1, 62) = 10.99$, $p = 0.002$]. The MLME model revealed that the (negative) combined Hedges' $g$ effect size estimates significantly deviated from zero in both studies in which listeners had to identify linguistic prosody [–1.38, SE = 0.21, 95% CI (–1.81, –0.96); $t(62) = -6.47$, $p < 0.001$] and studies in which listeners had to identify emotional prosody [–2.39, SE = 0.25, 95% CI (–2.88, –1.90); $t(62) = -9.77$, $p < 0.001$]. Furthermore, the combined Hedges' $g$ effect size was significantly smaller for studies where listeners had to identify linguistic prosody than for studies where listeners had to identify emotional prosody [1.01, SE = 0.30, 95% CI (0.40, 1.61); $t(62) = -3.32$, $p = 0.002$]. This indicates that CI users and simulation listeners identify both linguistic and emotional prosody less accurately than NH listeners and that electric hearing has an even bigger negative effect on the identification of emotional prosody than on the identification of linguistic prosody.

**Effect of $f_0$** • The omnibus test of the MLME model (k = 64) with only the factor $f_0$ confirmed that the effect size estimate was moderated by $f_0$ of the stimuli the listeners had to identify [$F(1, 62) = 20.50$, $p < 0.001$]. The MLME revealed that the combined Hedges' $g$ effect size estimates significantly deviated from zero in studies where $f_0$ of the stimuli listeners had to identify was unmanipulated [–1.89, SE = 0.19, 95% CI (–2.27, –1.50); $t(62) = -9.79$, $p < 0.001$] but not for studies where $f_0$ of the stimuli listeners had to identify was normalized [0.13, SE = 0.47, 95% CI (–0.81, 1.07); $t(62) = 0.28$, $p = 0.779$], where normalization was either done by flattening the $f_0$ contour (Marx et al. 2015, study no. 12.2) or by using a noninformative $f_0$ contour (van de Velde et al. 2017, study no. 26.3 and study no. 26.6). Furthermore, the effect size was significantly smaller for studies where $f_0$ of the stimuli listeners had to identify was normalized than for the studies where $f_0$ of the stimuli listeners had to identify was unmanipulated [2.02, SE = 0.45, 95% CI (1.13, 2.91); $t(62) = 4.53$, $p < 0.001$]. This indicates that CI users and simulation listeners identify linguistic and emotional prosody less accurately than NH listeners when $f_0$ of the stimuli is unmanipulated, but that when $f_0$ of the stimuli is normalized, the accuracy of identification does not significantly differ between electric and acoustic hearing.

# DISCUSSION

The meta-analysis reported in this article quantitatively demonstrates that CIs and vocoder simulations have a large and significant negative effect on the identification of prosody, where CI users and simulation listeners are less accurate in correctly identifying linguistic and emotional prosody in nontonal languages than NH listeners. This negative effect of electric hearing is consistent with previous research that has shown that, compared with NH listeners, CI users and simulation listeners are less accurate in correctly identifying various elements of speech, such as segmental elements of speech, speech in noise, speaker gender, lexical tone, as well as prosody (e.g., Shannon 2002; Shannon et al. 2004; Green et al. 2005; Pisoni 2005; Meister et al. 2011a,b; Fuller et al. 2014b; Başkent et al. 2016b; Plack 2018). Moreover, the analysis revealed that the effect size estimate was not moderated by group and thus that the effect of electric hearing did not significantly differ between CI users and simulation listeners. That is, the effect size estimate for studies comparing prosody identification accuracy of CI users with NH listeners did not significantly differ from the effect size estimate for studies comparing the accuracy of simulation listeners with NH listeners. This suggests that (averaging over vocoder parameters and CI settings) the vocoder simulations of the vocoder studies influenced prosody identification similarly to how the CIs of the CI studies influenced prosody identification. Moreover, given the comparable effect of electric hearing between CI-user studies and simulation-listener studies, the prosody identification deficiency can mainly be attributed to the electric hearing signal of CIs and CI simulations; any differences that can be identified between these listener groups did not significantly influence the effect of electric hearing. On the one hand, it could be argued that CI users may have a disadvantage with speech identification as they are affected by physiological changes resulting from hearing loss, leading to differences in the functioning of the auditory and language processing systems between CI users and simulation listeners. On the other hand, it could be argued that CI users may have an advantage over simulation listeners as they may be more accustomed to accommodating the low quality of transmission of the acoustic cues (Pisoni 2005; Başkent et al. 2016a,b; Wagner et al. 2019). Yet, despite all these differences between the listener groups, the end-result of the specific effect of electric hearing studied here (i.e., prosody identification, not taking nonauditory factors into account) was comparable. This finding provides quantitative evidence in support of the methodological suitability of vocoder simulations in these studies investigating the effect of electric hearing. Vocoder simulations with carefully selected parameters thus could provide an adequate estimate of the effect of CIs on prosody identification and may therefore be regarded as an appropriate methodological tool for the investigation of the influence of CIs on prosody identification in the absence of CI users.

Furthermore, results revealed that the effect of electric hearing may be modulated by prosody type. Namely, there was a significant difference in the effect of electric hearing depending on whether the CI users or simulation listeners were identifying linguistic prosody or emotional prosody; the identification of emotional prosody was more strongly compromised by CIs and vocoder simulations than linguistic prosody. One possible explanation for this effect is the difference in cue-weighting strategies between these two prosody types. A recent study has shown that both the NH and simulation listeners of that study attached the heaviest weight to the cue $f_0$ when identifying emotional prosody, but that they attached an equal amount of weight to the cues $f_0$ and duration for the identification of linguistic prosody (van de Velde et al. 2015). It is therefore unsurprising that the identification of emotional prosody has a bigger negative effect of electric hearing than the identification of linguistic prosody, because $f_0$ is degraded in the electric hearing signal of CIs and vocoder simulations, whereas the cue duration is not degraded (Moore 2003; Drennan & Rubinstein 2008; Meister et al. 2011). It should be noted that, despite the significantly larger negative effect of electric hearing on the identification of emotional prosody compared with linguistic prosody, the results of this meta-analysis showed that CIs and vocoder simulations have a large and significant negative effect on both the identification of linguistic prosody and on the identification of emotional prosody. The identification of prosody, for both prosody types, is thus strongly compromised by electric hearing via CIs and vocoder simulations.

For a number of studies included in the meta-analysis (k = 13) the stimuli were manipulated, whereas in the other studies (k = 51) the stimuli were unmanipulated. This distinction in itself did not lead to a significant difference in effect sizes. However, as it differed between studies which cue (i.e., $f_0$, intensity, or duration) was manipulated and whether the manipulations were applied to only one or to multiple cues, the influence of stimuli manipulations on prosody identification for each acoustic cue individually was also analyzed. This analysis revealed that the manipulation distinction (i.e., unmanipulated vs. normalized) does lead to significant differences in effect sizes for $f_0$, but not for intensity or duration. Focusing on $f_0$, a large and significant negative effect of electric hearing on the identification of prosody was found when $f_0$ of the stimuli was unmanipulated. Yet, when $f_0$ was normalized (i.e., by flattening the $f_0$ contour or by using a noninformative $f_0$ contour), electric hearing via CIs and vocoder simulations had a nonsignificant, small positive effect on the identification of prosody. As it happens, the three studies in which $f_0$ of the stimuli was normalized were the only studies included in our meta-analysis that yielded a positive effect of electric hearing (Marx et al. 2015, study no. 12.2, which used flattened $f_0$ contours; van de Velde et al. 2017, study no. 26.3 and study no. 26.6, which used noninformative $f_0$ contours). Yet, it should be noted that this effect did not reach significance. As such, it can be inferred from these results that CI users and simulation listeners are approximately equally accurate in correctly identifying linguistic and emotional prosody as NH listeners when the cue $f_0$ is normalized. A finding that can be explained by the fact that the degradation in fine spectrotemporal detail of speech transmitted by CIs and vocoder simulations result in a low quality of transmission of $f_0$ and $f_0$ is thus degraded in the electric hearing signal (Moore 2003; Drennan & Rubinstein 2008; Meister et al. 2011b). When $f_0$ is normalized, the cue is also unavailable to NH listeners, thus providing them with the same acoustic information as CI users and simulation listeners, resulting in comparable identification accuracy scores.

## CONCLUSION

This meta-analysis revealed a robust negative effect of electric hearing for the identification of linguistic and emotional prosody. This effect did not differ between CI studies and vocoder studies, providing quantitative evidence for the methodological suitability of vocoder simulations in studies on the effect of electric hearing. The analysis identified $f_0$ as the cue with the biggest influence on the effect of electric hearing. It can be concluded from this analysis that prosody identification is strongly compromised by electric hearing and that this is mainly due to the degradation of $f_0$ in the electric hearing signal. Improvement of prosody identification for this target group can possibly be achieved by enhancing the quality of transmission of $f_0$ cues in CIs, for example, by improving device or electrode design that can overcome limitations of auditory nerve stimulation, or by training CI users and simulation listeners to adjust their cue-weighting strategies to more effectively perceive degraded $f_0$ cues and/or to make better use of other and more reliable cues.

## ACKNOWLEDGMENTS

The authors thank Monita Chatterjee, Andrew Faulkner, Christina Fuller, Steven Gilbers, Johan Hanekom, Yi Ting Huang, Xin Luo, Hartmut Meister, Lucas Mens, Narges Moein, David Morris, Takayuki Nakata, Rosemary O'Halpin, and Louise Straatman for providing access to the statistical data of their studies to facilitate their inclusion in the meta-analysis.

This research was supported by the Center for Language and Cognition Groningen and by a VICI Grant (No. 918-17-603) from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw).

All authors of the present paper contributed to this work. M.K.E. designed and performed the meta-analysis and wrote the paper. All authors discussed the results and implications, commented on the manuscript at all stages, and approved the final version.

The authors have no conflicts of interest to disclose.

Received June 7, 2019; accepted January 20, 2020.

## REFERENCES


References marked with an asterisk indicate studies included in the meta-analysis.

*Agrawal, D., Timm, L., Viola, F. C., et al. (2012). ERP evidence for the recognition of emotional prosody through simulated cochlear implant strategies. *BMC Neurosci, 13*, 113.

Aguinis, H., Gottfredson, R. K., Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Org Res Methods*, *16*, 270–301.

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *Quant Methods Psychol*, *12*, 154–174.

Başkent, D., Clarke, J., Pals, C., et al. (2016a). Cognitive compensation of speech perception with hearing impairment, cochlear implants, and aging: How and to what degree can it be achieved? *Trends Hear*, *20*, 1–16.

Başkent, D., Gaudrain, E., Tamati, T. N., et al. (2016b). 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, Modeling, and Medicine* (pp. 285–319). Plural Publishing Inc.

Belyk, M., & Brown, S. (2014). Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies. *Soc Cogn Affect Neurosci, 9*, 1395–1403.

Bhargava, P., Gaudrain, E., Başkent, D. (2016). The intelligibility of interrupted speech: Cochlear implant users and normal hearing listeners. *J Assoc Res Otolaryngol, 17*, 475–491.

Blamey, P., Artieres, F., Başkent, D., et al. (2013). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients. *Audiol Neurootol, 18*, 36–47.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., et al. (2009). *Introduction to Meta-Analysis*. Hoboken: Wiley.

*Chatterjee, M., Zion, D. J., Deroche, M. L., et al. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hear Res, 322*, 151–162.

Cheung, M. W. -L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychol Methods*, *19*, 211–229.

Cheung, M. W. -L. (2015). *Meta-Analysis: A Structural Equation Modeling Approach*. Chichester: John Wiley & Sons, Inc.

Cumming, G. (2012). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. New York: Routledge.

Drennan, W. R., & Rubinstein, J. T. (2008). Music perception in cochlear implant users and its relationship with psychophysical capabilities. *J Rehabil Res Dev, 45*, 779–789.

Dunn, C. C., Walker, E. A., Oleson, J., et al. (2014). Longitudinal speech perception and language performance in pediatric cochlear implant users: The effect of age at implantation. *Ear Hear, 35*, 148–160.

Egger, M., Davey Smith, G., Schneider, M., et al. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*, 629–634.

*Fengler, I., Nava, E., Villwock, A. K., et al. (2017). Multisensory emotion perception in congenitally, early, and late deaf CI users. *PLoS One, 12*, e0185821.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *Br J Math Stat Psychol, 63*(Pt 3), 665–694.

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. *J Assoc Res Otolaryngol, 5*, 253–260.

Fu, Q. J., Chinchilla, S., Nogaki, G., et al. (2005). Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. *J Acoust Soc Am, 118*(3 Pt 1), 1711–1718.

*Fuller, C. D., Galvin, J. J. III, Maat, B., et al. (2014a). The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations? *Front Neurosci, 8*, 179.

Fuller, C. D., Gaudrain, E., Clarke, J. N., et al. (2014b). Gender categorization is abnormal in cochlear implant users. *J Assoc Res Otolaryngol, 15*, 1037–1048.

*Geers, A. E., Davidson, L. S., Uchanski, R. M., et al. (2013). Interdependence of linguistic and indexical speech perception skills in school-age children with early cochlear implantation. *Ear Hear, 34*, 562–574.

*Gilbers, S., Fuller, C., Gilbers, D., et al. (2015). Normal-hearing listeners' and cochlear implant users' perception of pitch cues in emotional speech. *Iperception, 6*, 0301006615599139.

Goedemans, R. W. N., Heinz, J., van der Hulst, H. (Eds.). (2018). *The Study of Word Stress and Accent: Theories, Methods, and Data*. Cambridge University Press.

Green, T., Faulkner, A., Rosen, S., et al. (2005). Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification. *J Acoust Soc Am, 118*, 375–385.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Hopewell, S., Clarke, M., Mallett, S. (2005). Grey literature and systematic reviews. In H. R. Rothstein, A. J. Sutton, M. Borenstein (Eds.), *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustment* (pp. 49–72). John Wiley & Sons Ltd.

*Hopyan-Misakyan, T. M., Gordon, K. A., Dennis, M., et al. (2009). Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. *Child Neuropsychol, 15*, 136–146.

*Huang, Y. T., Newman, R. S., Catalano, A., et al. (2017). Using prosody to infer discourse prominence in cochlear-implant users and normal-hearing listeners. *Cognition, 166*, 184–200.

*Kalathottukaren, R. T., Purdy, S. C., Ballard, E. (2015). Prosody perception and musical pitch discrimination in adults using cochlear implants. *Int J Audiol, 54*, 444–452.

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Stat Med, 22*, 2693–2710.

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Res Synth Methods, 2*, 61–76.

Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge: Cambridge University Press.

Ladefoged, P., & Johnson, K. (2010). *A Course in Phonetics* (6th ed.). Boston: Cencage Learning.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Front Psychol, 4*, 863.

Lee, K. Y., van Hasselt, C. A., Chiu, S. N., et al. (2002). Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children. *Int J Pediatr Otorhinolaryngol, 63*, 137–147.

Luo, X., & Fu, Q. J. (2009). Concurrent-vowel and tone recognitions in acoustic and simulated electric hearing. *J Acoust Soc Am, 125*, 3223–3233.

*Luo, X., Fu, Q. J., Galvin, J. J. III. (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends Amplif, 11*, 301–315.

Luo, X., Fu, Q. J., Wu, H. P., et al. (2009). Concurrent-vowel and tone recognition by Mandarin-speaking cochlear implant users. *Hear Res, 256*, 75–84.

*Lyxell, B., Wass, M., Sahlén, B., et al. (2009). Cognitive development, reading and prosodic skills in children with cochlear implants. *Scand J Psychol, 50*, 463–474.

*Marx, M., James, C., Foxton, J., et al. (2015). Speech prosody perception in cochlear implant users with and without residual hearing. *Ear Hear, 36*, 239–248.

*Meister, H., Landwehr, M., Pyschny, V., et al. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *Int J Audiol, 48*, 38–48.

Meister, H., Landwehr, M., Pyschny, V., et al. (2011a). Use of intonation contours for speech recognition in noise by cochlear implant recipients. *J Acoust Soc Am, 129*, EL204–EL209.

*Meister, H., Landwehr, M., Pyschny, V., et al. (2011b). The perception of sentence stress in cochlear implant recipients. *Ear Hear, 32*, 459–467.

*Moein, N., Khoddami, S. M., Shahbodaghi, M. R. (2017). A comparison of speech intonation production and perception abilities of Farsi speaking cochlear implanted and normal hearing children. *Int J Pediatr Otorhinolaryngol, 101*, 1–6.

Moore, B. C. (2003). Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants. *Otol Neurotol, 24*, 243–254.

*Morris, D., Magnusson, L., Faulkner, A., et al. (2013). Identification of vowel length, word stress, and compound words and phrases by postlingually deafened cochlear implant listeners. *J Am Acad Audiol, 24*, 879–890.

*Most, T., & Aviner, C. (2009). Auditory, visual, and auditory-visual perception of emotions by individuals with cochlear implants, hearing AIDS, and normal hearing. *J Deaf Stud Deaf Educ, 14*, 449–464.

*Nakata, T., Trehub, S. E., Kanda, Y. (2012). Effect of cochlear implants on children's perception and production of speech prosody. *J Acoust Soc Am, 131*, 1307–1314.

*O'Halpin, R. (2010). *The perception and production of stress and intonation by children with cochlear implants* (Doctoral dissertation). University College London, London, UK.

*Peng, S. C., Chatterjee, M., Lu, N. (2012). Acoustic cue integration in speech intonation recognition with cochlear implants. *Trends Amplif, 16*, 67–82.

Peng, S. C., Lu, H. P., Lu, N., et al. (2017). Processing of acoustic cues in lexical-tone identification by pediatric cochlear-implant recipients. *J Speech Lang Hear Res, 60*, 1223–1235.

*Peng, S. C., Tomblin, J. B., Turner, C. W. (2008). Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *Ear Hear, 29*, 336–351.

*Petersen, B., Mortensen, M. V., Hansen, M., et al. (2012). Singing in the key of life: A study on effects of musical ear training after cochlear implantation. *Psychomusicol: Music Mind Brain*, *22*, 134–151.

Pisoni, D. B. (2005). Speech perception in deaf children with cochlear implants. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 494–523). Wiley-Blackwell.

Plack, C. J. (2018). *The Sense of Hearing* (3rd ed.). New York: Routledge.

Raithel, V., & Hielscher-Fastabend, M. (2004). Emotional and linguistic perception of prosody: Reception of prosody. *Folia Phoniatr Logop, 56*, 7–13.

*See, R. L., Driscoll, V. D., Gfeller, K., et al. (2013). Speech intonation and melodic contour recognition in children with cochlear implants and with normal hearing. *Otol Neurotol, 34*, 490–498.

Shannon, R. V. (2002). The relative importance of amplitude, temporal, and spectral cues for cochlear implant processor design. *Am J Audiol, 11*, 124–127.

Shannon, R. V., Fu, Q.-J., Galvin III, J. J., et al. (2004). Speech perception with cochlear implants. In F.-G. Zeng, A. N. Popper, & R. R. Fay (Eds.), *Cochlear Implants: Auditory Prostheses and Electric Hearing*. Springer.

Sharma, A., Dorman, M. F., Kral, A. (2005). The influence of a sensitive period on central auditory development in children with unilateral and bilateral cochlear implants. *Hear Res, 203*, 134–143.

Sterne, J. A. C., Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, M. Borenstein (Eds.), *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustment* (pp. 99–110). John Wiley & Sons Ltd.

Stevens, J. P. (1984). Outliers and influential data points in regression analysis. *Psychol Bull*, *95*, 334–344.

*Straatman, L. V., Rietveld, A. C., Beijen, J., et al. (2010). Advantage of bimodal fitting in prosody perception for children using a cochlear implant and a hearing aid. *J Acoust Soc Am, 128*, 1884–1895.

Thompson, S. G., & Higgins, J. P. (2002). How should meta-regression analyses be undertaken and interpreted? *Stat Med, 21*, 1559–1573.

Tobey, E. A., Thal, D., Niparko, J. K., et al.; CDaCI Investigative Team. (2013). Influence of implantation age on school-age language performance in pediatric cochlear implant users. *Int J Audiol, 52*, 219–229.

*Torppa, R., Faulkner, A., Huotilainen, M., et al. (2014). The perception of prosody and associated auditory cues in early-implanted children: The role of auditory working memory and musical activities. *Int J Audiol, 53*, 182–191.

Torppa, R., Faulkner, A., Järvikivi, J., et al. (2010). Acquisition of focus by normal hearing and cochlear implanted children: The role of music experience. Paper presented at the *5th International Conference on Speech Prosody*, Chicago, IL.

van de Velde, D. J. (2017). *The processing of Dutch prosody with cochlear implants and vocoder simulations* (Doctoral dissertation). Leiden University, Leiden, The Netherlands.

van de Velde, D. J., Khoshchin, A., ter Beek, L., et al. (2015). Discrimination of emotional and linguistic prosody with cochlear implant simulations. Paper presented at the *18th International Congress of Phonetic Sciences (ICPhS)*, Glasgow, UK.

*van de Velde, D. J., Schiller, N. O., van Heuven, V. J., et al. (2017). The perception of emotion and focus prosody with varying acoustic cues in cochlear implant simulations with varying filter slopes. *J Acoust Soc Am*, *141*(5), 3349–3363.

*van Zyl, M. (2014). *Perception of prosody by cochlear implant recipients* (Doctoral dissertation). University of Pretoria, Pretoria, South Africa.

*Van Zyl, M., & Hanekom, J. J. (2013). Perception of vowels and prosody by cochlear implant recipients in noise. *J Commun Disord, 46*, 449–464.

Viechtbauer, W. (2010a). Conducting meta-analysis in R with the metafor pacakage. *J Stat Softw*, *36*(3), 1–48.

Viechtbauer, W. (2010b). metafor: Conduction meta-analyses in R with the metafor package (Version 2.1-0) [R package]. Retrieved October 31, 2019 from https://cran.r-project.org/web/packages/metafor

Viechtbauer, W., & Cheung, M. W. (2010). Outlier and influence diagnostics for meta-analysis. *Res Synth Methods, 1*, 112–125.

Wagner, A. E., Nagels, L., Toffanin, P., et al. (2019). Individual variations in effort: Assessing pupillometry for the hearing impaired. *Trends Hear, 23*, 2331216519845596.

Wang, W., Zhou, N., Xu, L. (2011). Musical pitch and lexical tone perception with cochlear implants. *Int J Audiol, 50*, 270–278.