Universiteit Leiden
The Netherlands

# The processing of Dutch prosody with cochlear implants and vocoder simulations

Velde, D.J. van de

**Citation**

Velde, D. J. van de. (2017, July 5). *The processing of Dutch prosody with cochlear implants and vocoder simulations*. *LOT dissertation series*. LOT, Utrecht. Retrieved from https://hdl.handle.net/1887/50406

| | |
|---|---|
| Version: | Not Applicable (or Unknown) |
| License: | Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden |
| Downloaded from: | https://hdl.handle.net/1887/50406 |

**Note:** To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

Leiden University Repository

The handle http://hdl.handle.net/1887/50406 holds various files of this Leiden University dissertation.

**Author**: Velde, D.J. van de
**Title**: The processing of Dutch prosody with cochlear implants and vocoder simulations
**Issue Date**: 2017-07-05

Chapter **6**

# Prosody perception and production by children with cochlear implants


## Abstract

Cochlear implant (CI) users have been reported to have difficulty perceiving and producing prosody. In this study the perception and production of emotional and linguistic (focus) prosody were compared in children with CIs and normally hearing (NH) peers.

Thirteen CI and Thirteen hearing-age (HA) matched NH children (HAs between 3;8 and 9;5) performed, as baseline tests, non-verbal emotion understanding tests (for general emotional development), a non-word-word repetition test (for general linguistic development) and stimulus identification and naming tests (for basic task understanding). Main tests were verbal emotion (happy, sad) discrimination, verbal focus position (color or noun) discrimination in simple color + noun sentences, acted emotion production and focus production (elicitation of corrective focus). Accuracy scores were compared across groups and correlations between tests were computed. Emotion and focus productions were evaluated by a group of 10 adult Dutch listeners with normal hearing.

The focus perception test could not be analyzed. Scores for the two groups were comparable for all tests, except a lower score for the CI group in the Non-word repetition test. On the individual participant level, emotional prosody perception and production scores were weakly and moderately significantly correlated for CI children but uncorrelated for NH children. In both groups, emotion production, but not emotion perception, was weakly predicted by hearing age. Non-verbal emotion (but not linguistic) prosody understanding performance, predicted CI children's emotion perception and production scores, but not the controls'.

Given the comparable overall scores, CI children catch-up with their peers no later than towards the end of primary school. Increasing time in sound facilitates vocal emotional expression, which possibly requires independently maturing emotion perception skills. CI and NH children apply the same cue-weighting strategies for emotion perception, relying almost exclusively on F0 information.

## 6.1 Introduction

Children with cochlear implants (CI) experience delays or deviations in their oral (productive and perceptual) linguistic and socio-emotional development relative to normally (NH) hearing peers. This is, first of all, because the onset of their oral language acquisition process is delayed until the moment of implantation (usually at least at one year of age). Second, due to the fact that the quality of the linguistic input that can be received after implantation is degraded compared to what NH peers can perceive, a full appreciation of phonetic nuances important for linguistic and paralinguistic information is hindered. For instance, CI users have been found to have problems with identifying vowels (Dorman & Loizou, 1998; Garrapa, 2014; Valimaa, Maatta, Lopponen & Sorri, 2002; Välimaa, Sorri, Laitakari, Sivonen & Muhli, 2011; however, see Iverson, Smith & Evans, 2006), distinguishing questions from statements (Meister, Landwehr, Pyschny, Walger & von Wedel, 2009; Peng, Lu & Chatterjee, 2009; Straatman, Rietveld, Beijen, Mylanus & Mens, 2010), understanding speech in noise (Gfeller, Turner, Oleson, Zhang, Gantz, Froman & Olszewski, 2007; Neuman, 2014), identifying emotions in speech (Geers, Davidson, Uchanski & Nicholas, 2013; Luo, Fu & Galvin, 2007) and discriminating speaker gender and identity (Fu, Chinchilla, Nogaki & Galvin, 2005; Fuller, Gaudrain, Clarke, Galvin, Fu, Free & Baskent, 2014; however, see Meister et al., 2009). Problems with the production of speech have also been observed, including voice quality (Ubrig, Goffi-Gomez, Weber, Menezes, Nemr, Tsuji & Tsuji, 2011), articulation (Van Lierde, Vinck, Baudonck, De Vel & Dhooge, 2005), lexical tone production (Han, Zhou, Li, Chen, Zhao & Xu, 2007), emotion imitation (Nakata, Trehub & Kanda, 2012; Wang, Trehub, Volkova & van Lieshout, 2013), intelligibility (Chin, Tsai & Gao, 2003), and the quality, content and efficiency of retold stories (Boons, De Raeve, Langereis, Peeraer, Wouters & van Wieringen, 2013). However, vocal characteristics within the norm have also been reported (Souza, Bevilacqua, Brasolotto & Coelho, 2012).

According to one series of studies testing 181 implanted children, speech perception and production performance have been shown to explain 42% of overall total language scores and as much as 63% when split for overall spoken language scores and (Geers, Nicholas & Sedey, 2003), showing the importance of speech perception and production for children’s linguistic development. Furthermore, problems in those areas have been associated with delays in socio-emotional development. Wiefferink, Rieffe, Ketelaar, De Raeve and Frijns (2013) tested Dutch CI and NH two-and-a-halve-to five-year-old children on facial and situational emotion understanding and general expressive and receptive language development. For the recipients, performance on all tests was poorer than for the control group and showed positive correlations between language and emotion tests that require verbal processing. These results showed that CI children experience delays in verbal as well as non-verbal emotion understanding and that linguistic development can predict aspects of emotional development. Mancini, Giallini, Prosperini, D’Alessandro H, Guerzoni, Murri, Cuda, Ruoppolo, De Vincentiis and Nicastri (2016), however, found that 79% of their cohort of 72 CI children, aged 4 to 11 years, showed normal emotion understanding skills. The differences with Wiefferink et al.’s results were attributed to discrepancies between the participant groups: Mancini et al.’s cohort had a wider age range and a larger percentage of children with an exclusively oral language use. It might be the case that CI children catch up for their delay in emotional development when they are at school age. Nevertheless, similarly to Wiefferink et al. (2013), Mancini et al. (2016) also reported a link between emotional and linguistic development of CI children.

One area of speech that has been relatively little studied in the research on the linguistic development of CI children and adults with CIs is prosody. Prosody is defined as the speech information which cannot be reduced to the individual segments (consonants and vowels) or their juxtaposition (Rietveld & Van Heuven, 2016). It is an essential component of speech because it conveys both messagerelated (meaning) and speaker-related (emotion and attitude) information. These types are referred to as linguistic and emotional prosody, respectively. For a number of possible reasons, linguistic and emotional prosody may develop differently in a language learner. First of all, their neurolinguistic processing is most likely partly lateralized, with emotional prosody being associated mostly with the right hemisphere and linguistic prosody with both hemispheres (Witteman, van Ijzendoorn, van de Velde, van Heuven & Schiller, 2011); second, they are phonetically different (linguistic information is discrete whereas emotional information is gradient); and third, the production of linguistic prosody plausibly requires knowledge of linguistic rules whereas that of emotional prosody, being more intuitive, might not, and might thus depend less on perception.

Whereas comprehension of sentences by pediatric and adult CI users has been found to be relatively intact (e.g., Helms, Müller, Schön, Moser, Arnold, Janssen, Ramsden, Von Ilberg, Kiefer & Pfennigdorf, 1997), several aspects of the perception and production of linguistic and emotional prosody have proven more problematic. As for the perception of linguistic prosody, Meister, Tepeli, Wagner, Hess, Walger, von Wedel and Lang-Roth (2007) reported poorer performance for adult CI users than for NH controls on the identification of word and sentence accent position and sentence type (question vs. statement), but not on discrimination of durational minimal pairs of words, and sentential phrasing with any available cue (e.g., *Die Oma schaukelt das Mädchen nicht.* vs. *Die Oma schaukelt. Das Mädchen nicht.*, lit. ‘Grandma swings the girl not’ vs. ‘Grandma swings. The girl not.’). Children with CIs were outperformed by peers with hearing aids (HA) in the discrimination of questions vs. statements and lexical stress position on bisyllables, but groups performed equally on the identification of words’ syllable number and sentence stress (narrow focus) position (Most & Peled, 2007). O’Halpin (2009) found lower performance for school-going children than for NH peers for phrasal discrimination (*blue bottle* vs. *bluebottle*) and identification of two-way (*It’s a BLUE book* vs. *It’s a* *blue BOOK*, where capitals demark accent) and three-way sentence accent position (*The BOY is painting a boat* vs. *The boy is PAINTING a boat* vs. *The boy is painting a BOAT*). Combined, these studies suggest that CI users have difficulty perceiving some but not all aspects of linguistic prosody, with a notable disadvantage for the identification of the position of accents on syllables and words (for evidence for similar difficulties by NH adults, see Schiller, 2006).

As for emotional prosody perception, Volkova, Trehub, Schellenberg, Papsin and Gordon (2013) found that five- to seven-year old implanted children discriminated happy and sad utterances with a score above chance but less accurately than NH peers. Children with CIs aged between seven and thirteen years in Hopyan-Misakyan, Gordon, Dennis and Papsin (2009) performed worse than NH peers when identifying the emotion (happy, angry, sad, fearful) of emotionally pronounced variants of semantically neutral sentences but the two groups performed equally on affective facial recognition, showing that difficulties with vocal emotion recognition could not be explained by more general delays in emotion understanding. In a study by Luo et al. (2007), adult recipients' scores were poorer than those of a NH control group when identifying the emotion (happy, angry, sad, fearful or neutral) of sentences. These studies show that CI recipients of various ages have difficulty identifying emotions in speech.

The main phonetic dimensions by which prosodic information is conveyed – dynamic, temporal and intonational (F0, fundamental frequency) variation – have been investigated to explain the mechanism behind CI users' prosody perception capabilities. Meister, Landwehr, Pyschny, Wagner and Walger (2011) measured difference limens (DL) and incrementally manipulated the F0, intensity and duration of accented syllables. They found that CI users had difficulty when F0 and intensity cues were made available but not when duration was made available, indicating that duration was more reliable for them than the other cues. These results were consistent with the findings that DLs for duration were comparable between groups (51 ms for CI vs. 40 ms for NH) but worse for the recipients for F0 (5.8 vs. 1.5 semitones) and for intensity (3.9 dB vs. 1.8 dB). The CI children in O'Halpin (2009) showed larger DLs than the control group in detection of F0 manipulated *baba* bisyllables but less so for intensity and duration. The variation in their performance was however large, with some participants showing smaller DLs than the smallest of the control group for intensity and duration. DLs per cue correlated with performance on the perception of phrasal accents reviewed above, which suggests that the children apply their successful psychophysical capabilities for prosodic perception. Taken together, it can be concluded from this research that CI users have problems discriminating variation in the intonational domain, but less in the dynamic and probably even less in the temporal domain and that this has repercussions for the type of prosodic information that they adequately receive.

A small number of studies have addressed the issue of prosody production by CI users. Lyxell, Wass, Sahlen, Samuelsson, Asker-Arnason, Ibertsson, Maki-Torkko, Larsby and Hallgren (2009) observed poorer performance for school-going CI children than for NH peers on the perception and production of word and phrase level prosody, but did not fully specify the task and phonetic analysis of the recorded data. Japanese children with CIs aged 5 to 13 years produced less appropriate imitations of disappointed and surprised utterances than a NH control group and their performance pattern was correlated to their impaired identification of emotions (i.e., happy, sad or angry) in semantically neutral sentences (Nakata et al., 2012). A below-normal performance but no correlation was found for six- to ten-year-old recipients between the Beginner's Intelligibility Test, a sentence imitation test for CI users (Osberger, 1994), and the Prosodic Utterance Production test, an imitation test for sentences with happy, sad, interrogative and declarative moods (Bergeson & Chin, 2008). Phonetic differences between CI relative to NH children's productions were found such as inadequate speech rate (longer utterances, longer pauses and schwas, more breath groups), inappropriate stress production and vocal resonance quality, a smaller F0 range and a shallower F0 declination, i.e., the natural downward F0 slope over an utterance (Clark, 2007; Lenden & Flipsen, 2007). Relative to NH peers, declarations and question produced by implanted children and young adults were less accurately identified as such (74% vs. 97%) and rated as less appropriate (3.1 vs. 4.5 on a scale from 1 to 5) by NH raters (Peng, Tomblin & Turner, 2008). In her study on school-going recipients, O'Halpin (2009) reported no correlation between most of the perception scores and production appropriateness of narrow focus position. The CI children in Holt (2013) produced phrasal emphasis (focus) sometimes with different accent types in terms of the autosegmental framework (Gussenhoven, 2004; Pierrehumbert, 1980) and with different syllabic alignments and temporal phrasing. In as far as they were able to produce the accents correctly, however, they did this without being able to discriminate between the accent types according to perception experiments, suggesting that accurate perception is not a prerequisite for reasonable production. In conclusion, as for perception, the production of both linguistic and emotional prosody by CI users of different ages deviates from the NH norm in several aspects. There is, however, mixed evidence regarding the question if good perception skills are required for good production skills.

The current research aimed at filling in this gap by testing the perception and production of linguistic and emotional prosody in the same group of implanted children and compare them to a control group of NH peers. The processing of linguistic and emotional prosody by implanted children has never been clearly contrasted. This line of research needs to be undertaken because the perceptional capabilities of CI children may have different repercussions for both the perception and production of the two types of prosody. Whereas the perception of both types may be affected by the degraded input (be it in a different manner or to a different degree), the production of emotional prosody is expected to be less affected than that of linguistic prosody due to its relatively intuitive, less rule-based nature.

In order to control for a number of known possible confounds, information about general linguistic level, emotion understanding and the family's socio-economic status was also gathered. We tested the following predictions.

> (A1) Prosody perception and production scores within participants are correlated. Such an effect would suggest that reasonable production skills require reasonable perception skills for a comparable task. (A2) That effect is larger within than across the prosody type (linguistic vs. emotional) and (A3) larger for linguistic than for emotional prosody because emotional production, due its supposedly relatively intuitive and less rule-based nature, is expected to be less dependent on perception skills.
>
> (B1) Scores per prosody type (linguistic or emotional) are influenced by their respective general scores for linguistic and emotional capacities, (B2) but this effect is larger for linguistic than for emotional prosody.
>
> (C1) Assuming a possible effect of more general maturation on linguistic, including prosodic, skills (hypothesis B), CI activation age negatively correlates with prosody processing capacities, but (C2) this effect is larger for linguistic than for emotional prosody.
>
> (D1) For the perception of prosody, CI participants rely more heavily on temporal cues as opposed to F0 cues than NH participants do. For NH participants, this reliance would be more equal between cues or the other way around. (D2) We expect that this effect is stronger for linguistic than for emotional prosody.

In summary, we investigated if scores on perception and production of prosody were related to each other per participant and if this relationship differed between linguistic and emotional prosody. We also studied to what extent these scores were related to more general linguistic and emotional capacities, and if CI users used different cues for prosody perception and production than the NH control group.

## 6.2 Methods

All children were tested on emotion perception, focus perception, emotion production and focus production, the order of which was randomized across participants. This block of four main tests was preceded by a familiarization phase, in which participants were acquainted with the names of the stimuli (colors and objects). Additionally, there were four baseline tests with the purpose of assessing the levels of possibly confounding competences: non-verbal emotion understanding, stimulus identification and naming, and non-word repetition, the first three of which took place before the main tests and the last of which after them, if the child's concentration allowed. The non-verbal emotional understanding comprised two tests from a battery designed to assess social-emotional development in normally hearing and children with special (linguistic) developmental or language backgrounds such as those with cochlear implants (Wiefferink, de Vries & Ketelaar, 2015). This test was included to ensure that all participants had a basic understanding of emotions, tested without the requirement of good verbal expression. All other tests were developed by the authors for the current research. The stimulus identification and naming tests were used as a baseline assessment of the capability to understand and name the stimuli to be used in the main tests. The non-word repetition test was included as a proxy for general linguistic capacities, which might or might not correlate with scores on tests gauging prosody processing capacities. The parents or caretakers were asked to complete a questionnaire about their socio-economic status (SES) and the child's linguistic and medical background. The study was approved by the Leiden University Medical Center's (LUMC) medical ethical committee (NL46040.058.13).

It should be noted at this point that, due to a technical error, no data for the focus perception test had been collected. The description of the methodology will therefore focus on the other tests.

### *6.2.1 Participants*

Thirteen implanted children and thirteen children with normal hearing (NH) participated in this study. They were matched on gender (in both groups eleven boys) and hearing age, defined as the time since the onset of stable hearing, which is implant activation date for recipients and the date of birth for controls. The CI group's mean hearing age was 6;10 (years;months) (ranging between 3;8 and 9;5 and with an SD of 1;9) and the NH group's mean hearing age was 6;9 (range: 4;5-9;4; SD: 1;6). The CI group's mean chronological age was 9;1 (range: 6;1-12;3; SD: 2;0) and that of the NH group was by definition identical to its hearing age. Chronological age is defined as the time since birth. We used the following inclusion criteria for participants (both CI and NH unless not applicable): at least three years gross of CI experience, unilateral implantation, no reported medical problems related to the CI, Dutch as the only first language, no attested psychosocial and (only NH) audiological or speech problems. NH children were not subjected to audiological testing since their hearing was supposed to be better than that of the CI children to begin with. Participant characteristics are shown in Table 1.

### *6.2.2 Stimuli*

Speech stimuli for all tests were recorded as natural utterances in an anechoic booth with a sampling rate of 44.100 Hz and a sampling depth of 16 bit and were pronounced by a child language acquisition expert (CL). She was asked to pronounce stimuli at a regular pace and with specific prosody such that, where applicable, emotions and focused words would be clear for young children.

In the emotion perception test, all trials were based on six object names and six color names in Dutch: 'auto' (car), 'bal' (ball), 'ballon' (balloon), 'bloem' (flower), 'schoen' (shoe), 'stoel' (chair),

**Table 2.** Demographic and implant characteristics of CI recipients. Hearing age refers to the time since implantation. 'AB' is the Advanced Bionics HiRes 90k HiFocus 1j implant; 'Nucleus' is the Nucleus Freedom Contour Advance implant. Abbreviations: x;y – years;months; mos.: months.

| Subject number (gender) | Chro-nolog -ical age | Estimated age at hearing loss onset | Estimated duration of deafness (months) | Age at first CI active-tion | Hea-ring age (mos.) | Etiology | Im-plan -ted ear(s) | Current implant type | Current speech proces-ssor |
|---|---|---|---|---|---|---|---|---|---|
| 1 (M) | 10;1 | 0;0 | 11 | 0;11 | 109 | congenital, hereditary | Both | AB | Neptune |
| 2 (M) | 8;0 | 0;0 | 15 | 1;3 | 80 | unknown (sudden) | Both | AB | Neptune |
| 3 (M) | 11;10 | unknown | unknown | 8;1 | 44 | unknown | Right | AB | Neptune |
| 4 (F) | 8;2 | 0;0 | 13 | 1;1 | 84 | congenital | Right | AB | Neptune |
| 5 (M) | 12;3 | unknown | unknown | 4;10 | 88 | unknown | Left | AB | Neptune |
| 6 (M) | 10;7 | 0;3 | 9 | 1;2 | 113 | unknown (sudden) | Both | AB | Neptune |
| 7 (M) | 10;8 | unknown | unknown | 5;1 | 67 | unknown | Left | Nucleus | Cochlear CP810 |
| 8 (F) | 6;6 | 0;0 | 21 | 1;9 | 57 | Chudley Mc-Cullough | Left | AB | Neptune |
| 9 (M) | 8;1 | 0;0 | 14 | 1;2 | 83 | congenital | Both | AB | Neptune |
| 10 (M) | 10;10 | 0;0 | 21 | 1;9 | 109 | congenital | Both | AB | Neptune |
| 11 (M) | 6;1 | 0;0 | 11 | 0;11 | 61 | congenital | Both | AB | Neptune |
| 12 (M) | 8;1 | 0;0 | 14 | 1;2 | 83 | congenital, hereditary | Both | AB | Neptune |
| 13 (M) | 7;2 | 0;0 | 12 | 1;0 | 73 | congenital | Both | AB | Neptune |

'blauw' (blue), 'geel' (yellow), 'groen' (geen), and 'rood' (red). These words were chosen on the basis of a number of criteria: (1) they consisted mainly of voiced segments such that the intonation pattern would be least interrupted; (2) they were supposedly not semantically biased towards any emotion; (3) they had no inherent color bias, to avoid anomalies such as green bananas and blue trees; (4) nouns had common neuter, so they had the same article and adjectival declination; and (5) the nouns were known by at least 86% of children aged 2;3 years as tested by a questionnaire with 961 (pairs of) parents and listed in the Lexilijst (Schlichting & Lutje Spelberg, 2002). According to that questionnaire, the colors were known by between 47% and 63% of children of that age. However,  they were the four most frequent colors known by young children, our participants had a higher hearing age than 2;3 years and they were familiarized with the stimuli before the test phase. Words ending in voiceless segments were dispreferred because they interrupt the intonation contour but in our choice of stimuli priority was given to the criteria of familiarity and natural color-neutralness. Therefore, some voiceless segments are present in the list. Auditory stimuli had normalized amplitudes by scaling to peak (0.99). All stimuli were prerecorded because we wanted to prevent inter-token variation in the stimuli. They were presented in auditory-only modality to prevent clues from lip-reading, for which the experimental group might have an advantage.

In the Emotion perception test, all 24 combinations of the six objects and four colors were produced in a happy and a sad variant. The phrases followed the template 'een' [color] [N], where 'een' is the singular indefinite article. They were between 1.38 and 1.93 seconds long, with an average duration of 1.72 seconds for happy and 1.62 for sad phrases. It has been reported elsewhere (van de Velde, Schiller, van Heuven, van Ginkel, Briaire, Beers & Frijns, forthcoming) that the emotions, taken into account possible response biases, could be discriminated at near-ceiling level in the unprocessed condition by NH listeners, ensuring that the intended emotions and focus positions were successfully conveyed.

Sentences were all manipulated into three extra variants by cross-splicing aspects of the prosody from the non-neutral stimuli to the same neutral equivalents (the Cue condition): (1) only the F0 contour (F0 condition); (2) only the durations of the allophones (Duration condition); and (3) both the F0 contour and the allophone durations (Both condition). This was done in order to control the cues available to the participants. Because unique neutral variants (i.e., one single variant for the two emotions) constituted the bases of the stimuli, judgements by participants could only be based on F0, allophone durations, or both, respectively. Except for these cues, the two emotions were identical, since the underlying segmental material was identical for both emotion variants of a given phrase.

In all relevant tests, response options were represented with additional images. Pictures recurring in different tests were those depicting the auditory noun and color stimuli. They were based on the database of the Max Planck Institute in Nijmegen and were controlled for the number of pixels, name agreement, picture familiarity and age of acquisition for five- to six-year-old children (Cycowicz, Friedman, Rothstein & Snodgrass, 1997). These original line drawings were filled with basic colors using Microsoft Paint in order to be able to contrast colored objects with each other. All children were familiarized with the visual stimuli before testing by showing all color and object pictures as well as their combinations one by one and in groups, and inviting them to name them, the researcher correcting and asking to repeat whenever necessary. Pictures were controlled for the total number of pixels per picture.

In baseline test 1, Non-verbal emotion understanding, the stimuli and procedure in this test were developed by Wiefferink et al. (2015), to which we refer for details about stimuli. In the baseline tests 2 and 3, Stimulus identification and naming, the stimuli consisted of the auditory and visual materials that were also used in the four main tests, i.e., (subsets of) the 24 color/object combinations. The auditory stimuli were always the identical tokens of the same phrase and the visual materials the exact same pictures. In the emotion perception and production tests, there were, additionally, simple line drawings of a happy and a sad face.

In baseline test 4, the Non-word repetition test, stimuli consisted of nonsense words in a carrier phrase presented as the supposed words for phantasy toys of which colored photos accompanied the auditory stimuli. These photos were taken from a database developed for non-word repetition tests, designed to avoid associations with known objects or with emotions, particularly by children (Horst & Hout, 2015). The nonsense words were four stimuli of each word length from one to five syllables. They were based on De Bree, Rispens and Gerrits (2007), but adapted for children with a linguistic age of 3;0 years. The criteria for the phonological composition of the nonsense words, based on Dollaghan and Campbell (1998), were as follows: (1) they began and ended with consonants (Cs); (2) they contained no consonant clusters; (3) to ensure that non-word repetition would not be affected by a participant's vocabulary knowledge, non-words were constructed such that none of their individual syllables (CV or CVC) corresponded to a Dutch word; (4) they only contained phonemes that even atypically developing children with a chronological age of 2;8 years have acquired according to Beers (1995), and excluding the 'late eight' (i.e.consonants that are acquired late; Shriberg & Kwiatkowski, 1994), except for /s/ (which would have left too few possibilities to work with); (5) they contained only tense vowels, as these are perceptually more salient and less likely to being reduced to schwa than lax vowels; (6) to limit syllabic positional predictability, consonants, except /s/, occupied only positions in which they occurred less than 32% of their occurrences (Van Oostendorp, personal communication); (7) for independent recall of all consonants, they appeared only once in a word. Practice stimuli were different from the experimental stimuli. The carrier phrase of all non-words was the exact same token of 'Kijk! Een [word], een [word]. Kan jij dat zeggen?' ('Look! A [word], a [word], can you say that?'). The target words were spliced into the indicated slots. The complete lists of non-words can be found in Appendix C.

### *6.2.3 Procedure*

Testing took place in the children's homes, at the Leiden University phonetics laboratory or at the Leiden University Medical Center, depending on the parents' preference. Testing was divided over multiple sessions if time and concentration limits requested so. Combined visits had a duration of between one and two and a half hours. Testing started with the Non-verbal emotion understanding test and was followed by color/object identification and naming to familiarize the children with the stimuli and the paradigms at hand. Subsequently, we administered the four main tests, emotion and focus perception and production, in a counterbalanced order across participants. Finally, depending on time and motivation of the children, non-word repetition was tested. All tests except the non-verbal emotion understanding and stimulus identification and naming were preceded by practice stimuli that could be repeated if deemed necessary by the experimenter. All but the Non-verbal emotion understanding test were performed on a touchscreen computer. If the child pointed without touching, the experimenter selected the intended option for the child. There was no time limit for trials in any of the tests. The experimenter globally supervised the procedure throughout by explaining the tests and continuing to a next trial whenever this was not automatic. In all computer tests, the experimental part was preceded by a practice phase of between two and four trials, repeated maximally once when the experimenter thought the child did not understand the task well enough. In the practice phase, responses prompted feedback in the form of a happy or a doubtful smiley, all in greyscale to prevent biases towards any experimental color.

All tests except the Non-verbal emotion understanding test were run on a Lenovo 15 inch touchscreen laptop with the keyboard flipped backwards so children could easily reach the screen. Stimuli were played through a single Behringer MS16 speaker placed centrally over the screen. The distance from the speaker to the tip of the child's nose was set at 61.5 cm at zero degrees azimuth at the start of testing. Hardware settings were adapted for every participant to calibrate the sound level at 65 dBSPL at the ear using a Trotec BS 06 sound meter. This portable meter was calibrated to a high-quality A-weighted sound level meter on the basis of a one-minute steady stretch of noise with the same spectrum as that a large portion of the combined stimuli (thus from the same speaker) of the experiments. Note that the usage of headphones was not an option as they would interfere with children's implants. Presentation of auditory stimuli was mediated by a Roland UA 55 external sound card. In the prosody production and Non-word repetition tests, speech was recorded using a Sennheiser PC 131 microphone as input to a Cakewalk UA-1G USB audio interface. All computer tests were run with *E-Prime 2.0 Professional* (Psychology Software Tools, Pittsburgh, PA, USA;

Schneider, Eschman, & Zuccolotto, 2012) and *Powerpoint 2010* on a *Windows 8.1* operating system.

*Baseline test 1, Non-verbal emotion understanding.* This test consisted of the subtests Face discrimination, Face identification and Expression. The first involved sorting four series of eight line drawings into one of two categories: cars or faces, faces with or without glasses, faces with a negative (angry, sad) or positive (happy) emotion, and sad or angry faces, respectively. In the first and third series only, the first two trials were done by the experimenter as an example. In the second subtest, divided over two pages, there were two instances of line drawings of faces for each of the emotions happy (twice on one page), sad, angry fearful (twice on the other page). The child was asked to indicate consecutively which face showed each of these emotions, and, for each emotion, if another face showed that as well. In these two subtests, numbers of correct responses were recorded. In the third subtest, the child was presented with eight line drawings of emotion evoking situations (two of each of the emotions happy, sad, angry and scared) and was asked to tell how the protagonist, always shown from behind the head to avoid cues from the facial expression, felt, to match one of four emotional faces to it and to tell why the protagonist felt that way. In case he or she did not respond, each question was repeated once. The verbal and drawn emotion chosen were recorded as well as the verbatim response.

*Baseline tests 2 and 3, Stimulus identification and naming.* In the first of these two tests, stimulus identification, the child consecutively identified each of all of the 24 auditory object/color combinations by selecting a picture on screen. The target position was counterbalanced, as were the position and type of the distractors (only different color, only different object, both different). There was no time limit. Performance was calculated as percentage correct. Also, to prevent unnecessary proliferation of the number of trials, only six of the possible fifteen object contrasts were used, namely car-flower, ball-shoe, balloon-chair, flower-ball, shoe-car, chair-balloon (the first one being the target). These pairs were both conceptually and (in

Dutch) phonologically well distinctive. All objects in this shortlist functioned exactly once as a target and once as an object distractor. To make the task easy and to circumvent red-green color blindness, only two color contrasts were used, namely blue-red and green-yellow (twelve times each). In the second test, stimulus naming, subsequently, the same stimuli as in the identification test appeared as pictures on screen and the child was asked to name them as a color/object noun phrase (e.g., *Een rode bal*, ‘A red ball’) using the vocabulary from the identification test and trained for in the familiarization test. Responses were recorded as audio files and scored as accurate or inaccurate (wrong, unclear or no response), neglecting the presence or choice of a determiner.

*Baseline test 4, Non-word repetition*. This test consisted of twenty trials in series of four for each of the lengths from one to five syllables (four times five), consecutively. Children were asked to repeat the word they heard once. Responses were recorded to be scored later. Pictures and auditory stimuli for a trial were presented simultaneously. The picture remained visible until the next trial started.

*Main test 1, Emotion perception.* In this test, participants heard a phrase pronounced in either a happy or a sad manner. They were asked to indicate which emotion was conveyed by touching or pointing at the corresponding picture of an emotional face on the screen. There were three counterbalanced blocks of 24 randomized trials separated by breaks, differing in Cue and each preceded by two warm-up trials. A trial consisted of a fixation animation (1,250 ms), the stimulus presentation (indefinite time) and an inter-stimulus interval (ITI, 200 ms). During stimulus presentation, the two response options were shown on the screen to the left and right, as well as a depiction of the pronounced phrase (e.g., a blue ball). The response option positions were swapped halfway through the test for counterbalancing, which was indicated by an animation of the faces moving to their new position.

*Main test 2, Emotion production.* In this test, children were asked to act emotions using the words and emotion depicted. For instance, if they saw a picture of a red chair and a happy face, they were required to say ‘red chair’ in a happy way. Variants with different articles and plurals were accepted. There was no time limit for a trial. There were eight trials, namely two objects to be named with each of the emotions ‘happy’, ‘sad’, ‘angry’ and ‘scared’. There were no warm-up trials.

*Main test 4, Focus production.* The children verbally responded to prerecorded questions eliciting focus prosody. The questions of the form ‘Is this a [color] [N]?’ either matched (half of the stimuli) a picture they produced or contrasted in the color or in the noun (both a quarter of the stimuli). There were 24 stimuli on a single block, preceded by two warm-up trials. Trials were similar in setup to those of the emotion and focus perception tests.

### *6.2.4 Data analysis*

Group comparisons (CI vs. NH) were, when single values per participant were compared, performed with non-parametric tests because of the small sample sizes. A significance level of $p = 0.05$ was adopted. Analyses were performed using *SPSS version 23.0* (IBM Corp, Armonk, NY). Effect sizes are reported for two-way comparisons, as less fine-grained comparisons were not the endpoints of interest.

*Baseline test 1, Non-verbal emotion understanding.* In the Face discrimination and the Face identification tasks, the groups’ mean numbers of correct responses were computed and compared for all trials pooled together, using the Mann-Whitney *U* test for independent samples. In the Face discrimination task, this was done for all test components pooled as well as for each component separately, i.e., by addition of numbers of correct responses for both response options of an object or face pair (cars vs. flowers, faces with glasses vs. hats, faces with positive vs. negative expressions, and faces with sad vs. angry expressions). In the Expression task, mean response accuracy was compared between groups, separately for the verbal and the pointing responses. For both these response types, a distinction was made between strict and tolerant evaluation policies. In the strict policy, each trial was assigned one of four expected (prototypical) emotions (happy, angry, sad, scared) and a response counted as accurate if and only if that exact emotion was chosen. In tolerant policy, only a distinction between positive (happy) and negative (angry, sad, scared) emotions was made. Positive or negative vocabulary other than the expected emotion labels were tolerated as well. For both these policies, analyses were performed.

*Baseline tests 2 and 3, Stimulus identification and naming*. These data were analyzed by computing percentages correct. For the Stimulus identification test, this involved the percentage of accurately identified phrases by selecting the picture on the screen corresponding to the phrase heard. For the subsequent Naming test, this involved overtly naming the picture shown on screen using the vocabulary encountered in the Identification test. Responses were recorded to allow evaluation of naming accuracy (by the first author).

*Baseline test 4, Non-word repetition.* All responses were transcribed using broad IPA (International Phonetic Alphabet) transcription by the first author as well as, for a reliability check, 130 items (25%; equally drawn from all participants and as equally as possible from all items) by a trained Dutch phonologist unaware of the target pronunciations. Based on guidelines by Dollaghan and Campbell (1998), they were scored on a phoneme by phoneme basis, every omission or, contrary to Dollaghan and Campbell (1998), addition of a phoneme and substitution by another phoneme counting as an error. In case of omitted or added syllables, utterance were aligned with the target in such a way as to minimize the number of errors. Subsequently, the numbers of phonemes repeated correctly was divided by the total number of target phonemes per word yielding a Percentage of Phonemes Correct (PPC) per stimulus length in number of words (ranging between one and five) (Dollaghan & Campbell, 1998). These measures were compared between groups (CI and NH).

*Main tests 1, Emotion perception test.* Because in the Emotion test, only two response options were available, following Signal Detection Theory, scores were transformed into hitrates, with one value per subject per Phonetic Parameter (Stanislaw & Todorov, 1999). In this way, possible response biases were accounted for. Following Macmillan and Kaplan (1985), perfect scores for a subject in a cell, which are not computable, were replaced by 100%/2*N*, where *N* is the number of items in the cell (24). Results are presented as *d'* scores. Data were subsequently subjected to a Repeated Measures ANOVA, with Phonetic Parameter as the within-subjects variable and Group as the between-subjects variable.

*Main tests 2 and 3, Emotion production and Focus production.* Participants' verbal responses in the Emotion and Focus production tests were evaluated by a single panel of 10 Dutch adults with a mean age of 27.3 years who did not present a hearing loss of over 40 dBHL at any of the octave frequencies between 0.125 and 8 kHz, as audiometrically assessed (Audio Console 3.3.2, Inmedico A/S, Lystrup, Denmark). In the Emotion test, listeners judged by button-press which of four emotions (happy, angry, sad, scared) was conveyed independent of the contents of the utterance. In the Focus test, they judged which of three focus positions (color, object or both) was accented. Another condition of the Focus production, in which the question posed to the children corresponded in both color and object to the image displayed, was not further analyzed. In this test, listeners were explained the procedure of the production task and asked to imagine to which question the speaker's utterances were a response to, so that they would judge the phrasal accents as corrective focus realizations (which is how they were intended by the speakers). In both evaluation tests, the order of response options was counterbalanced between two different versions. The order of the two tests per listener was also counterbalanced.

For every trial, for each participant, ten correct or incorrect responses were considered, according to the evaluations by the panel of ten adult listeners. A child's production counted as correct when the emotion it was prompted to produce in the task corresponded to the emotion perceived by an adult listener, and counted as incorrect otherwise. This yielded 1,910 data points in the Emotion production test and 2,780 data points in the Focus production test. Percentages correct were calculated over this entire dataset and compared between Groups and Emotions. No *d'* scores were calculated, as is common for alternative forced choice (AFC) tasks with more than two options (Macmillan & Creelman, 2004).

## 6.3 Results

*Parent questionnaire.* Parents of NH children reported Dutch to be their own first language as well as the mother tongue and first language of their child, used at home, at school and with friends. One child had been treated for hearing problems and one other child had received speech therapy. No NH children had been treated by a neurologist or for social problems and none had problems with their sight. The average SES, computed as the sum of the questionnaire ranks of the two parents' highest level finished education and their income category, of this group was 19.4, ranging between 17 and 21 and with a SD of 1.6. Parents of CI children also reported Dutch as their first language. Their child's first linguistic input was reported as Sign Language of the Netherlands (SNL) received from parents who learned it as a second language or from Dutch Sign Language (DSL) teachers. Three parents indicated that the acquisition of Dutch was simultaneous with that of DSL and that two parents had not reported DSL's acquisition onset age. All parents of CI children indicated that communication with their child before implantation was more frequent (answers of three parents missing) and (except for two parents) easier using sign language and all of them reported (except one missing answer) that after implantation spoken language communication was more frequent and easier, showing that implantation had successfully given access to spoken language. One

CI recipient had been treated by a neurologist but no children had been treated for social problems. One CI recipient had problems with his/her sight. The average SES of this group's parents was 18.0, ranging between 12 and 22 and with a SD of 3.4.

*Baseline test 1, Non−verbal emotion understanding.* In the Face discrimination task, mean numbers of correct responses were not different between groups for all object or face pairs together ($U$ = 1230.5, $z$ = −1.17, $p$ = .24, $r$ = −.23) or any of the pairs separately according to Mann-Whitney $U$ tests (cars vs. flowers: $U$ = 84.5, $z$ = 0; $p$ = 1, $r$ = 0; faces with glasses vs. hats: $U$ = 71.5, $z$ = −1.44, $p$ = .51, $r$ = −.28; negative vs. positive faces: $U$ – 77.0, $z$ = −.56, $p$ = .72, $r$ = −.11; angry vs. sad faces: $U$ = 74.5, $z$ = −.56, $p$ = .61, $r$ = −.11; exact significance). In the Face identification task, no effect of group on the number of correct responses was found either ($U$ = 53.0, $z$ = −1.8, $p$ = .11, $r$ = −.35; exact significance). In the Expression task, no effect of group on mean accuracy scores was found for strict ($U$ = 4724.5, $z$ = −1.0, $p$ = .32, $r$ = −.20) and tolerant ($U$ = 4892.5, $z$ = −1.8, $p$ = .074, $r$ = −.35) verbal responses, nor for the strict ($U$ = 5267.5, $z$ = −.26, $p$ = .79, $r$ = −0.051) and tolerant ($U$ = 5253.0, $z$ = −1.4, $p$ = .16, $r$ = −.27) pointed responses. These results suggest that, to the degree tested, the two groups have comparable levels of non-verbal emotion understanding.

*Baseline tests 2 and 3, Stimulus identification and naming*. In the Identification test, the CI group scored 98.7% correct and the NH group a 100%. In the Naming test, CI group's accuracy was 100% and the NH group's accuracy 99.4%. There were no missing cases. These results show that both groups were sufficiently able to perform the kind of tasks that the main part of the study consisted of, namely identification and verbal responding. Moreover, the results show that subjects knew the words corresponding to the pictures used.

*Baseline test 4, Non-word repetition*. In the Non-word repetition test, 3 out of 520 productions (0.006%) were missing. The second rater's transcription of 20% of data corresponded for 93.8% to those by the first rater with disagreement occurring almost exclusively at the phonetic level of individual phonemes such as voicing, showing the first rater's transcription to be reliable. Of the remaining data, Table 2 and Figure 1 summarize the results, showing mean percentages of phonemes correct (i.e., correctly repeated) per group and per item length, in number of syllables. The two groups show a parallel downward pattern with increasing item length, but the CI recipients consistently show a lower score by around 5%. The relatively low percentages for the one- and two-syllable words is due to the relatively large percentage of mispronounced (or misheard) final nasal consonants in those words. The overall score was statistically significantly different between the two groups according to a *t*-test with equal variances not assumed ($t(1,515) = -3.2$, $p = .001$, $r = .69$). The NH group was therefore somewhat more accurate at repeating non-words than the CI group.

*Main test 1, Emotion perception.* Table 3 and Figure 2 show *d*' scores in the Emotion perception test, split by Phonetic parameter (Intonation, Temporal or Both) and subject group. Repeated measures ANOVA on the *d'* scores revealed a main effect of Phonetic parameter ($F(2,22) = 49.79$, $p < .001$), but no effect of Group ($F(1,23)$

**Table 2.** Mean percentage phonemes correct and standard deviations (in parentheses) correct per syllable length (in number of syllables) and per participant group (CI or NH) in the Non-word repetition test.

| Group | Mean Percentage Phonemes Correct (SD) | | | | | |
|---|---|---|---|---|---|---|
| | Item length (number of syllables) | | | | | |
| | 1 | 2 | 3 | 4 | 5 | **Total** |
| CI | 88.5 (17.3) | 87.3 (13.7) | 92.3 (12.2) | 85.0 (18.4) | 66.8 (23.5) | **84.0 (19.5)** |
| NH | 92.2 (14.3) | 91.5 (11.4) | 97.8 (7.7) | 89.3 (12.1) | 74.3 (21.2) | **89.0 (16.1)** |
| **Total** | **90.3 (15.9)** | **89.4 (12.8)** | **95.1 (10.5)** | **87.1 (15.7)** | **70.5 (22.6)** | **86.5 (18)** |

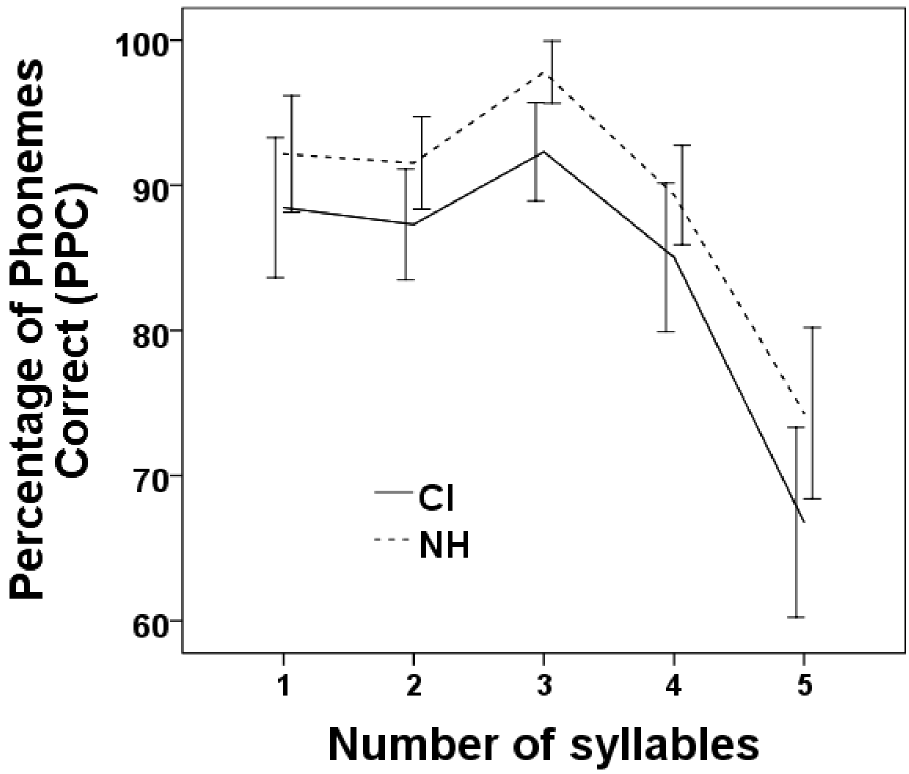


**Figure 1.** Percentage of Phonemes Correct per number of syllables in the Non-word repetition test. Percentages correct represent percentages of correctly repeated phonemes per non-word. Additions, omissions and substitutions of phonemes counted as errors.

= .18, $p$ = .68, $r$ = .39), nor an interaction between Phonetic parameter and Group ($F(2,22) = .29$, $p$ = .97). Post-hoc analyses revealed that of the three Phonetic parameters, scores on the Temporal condition differed highly significantly from both Intonation ($t(24) = 7.61$, $p < .001$, $r$ = .84) and Both ($t(25) = -10.70$, $p < .001$, $r$ = .91), but the Intonation and Both conditions were not significantly different from each other ($t(24) = -1.79$, $p$ = .086, $r$ = .34) given a Bonferroni-corrected $p$-criterion of .05/3. These results suggest that CI and NH groups were equally capable of discriminating the two emotions and that they do that applying the same cue weighting strategy.

**Table 3.** Mean *d'* scores split by Phonetic parameter and by participant group (CI or NH) in the Emotion perception test. Participants judged if prerecorded utterances were pronounced with a happy or sad emotion. Phonetic parameters indicate which type of phonetic information was available in the stimulus.

| Group | *d'* | | | |
|---|---|---|---|---|
| | Phonetic parameter | | | |
| | Intonation | Temporal | Both | Total |
| CI | 2,40 (1,26) | 0,52 (0,65) | 2,80 (1,05) | **1,91 (1,41)** |
| NH | 2,32 (1,19) | 0,26 (0,55) | 2,64 (1,2) | **1,72 (1,47)** |
| **Total** | **2,36 (1,2)** | **0,39 (0,61)** | **2,72 (1,11)** | **1,82 (1,43)** |

*Main test 2, Emotion production.* In the Emotion production test, of all trials, 3.8% were missing (missing response or technical error). Table 4 and Figure 3 show mean percentages correct of the four emotions in both participant groups (CI and NH). The overall accuracy of the CI group (62.3%) was somewhat higher than that of the NH group (57.8%) but the group difference varied across emotions. According to two- and four-way ANOVAs, respectively, there was a very small but significant effect of Group ($F(1,1902) = 7.06$, $p = .008$, $r = .061$) and one of Emotion ($F(3,1902) = 45.43$, $p < .001$), as well as a significant interaction between Group and Emotion ($F(3,1902) = 7.82$, $p < .001$). Bonferroni-corrected post-hoc tests showed that all levels of Emotions differed highly significantly ($p < .001$), except angry and sad ($p = 1$). Separate Group comparisons for each emotion showed that the CI group scored higher than the NH group on scared ($F(1,438) = 10.06$, $p = .002$, $r = .15$) and angry ($F(1,478) = 14.01$, $p < .001$, $r = .17$) responses, that the NH scored better on happy responses ($F(1,298) = 5.11$, $p = .024$, $r = .13$), but that there was no difference for sad responses ($F(1,488) = .017$, $p = .90$, $r = .019$). These results indicate that the two groups have different specialties when it comes to the production of emotions, but that in

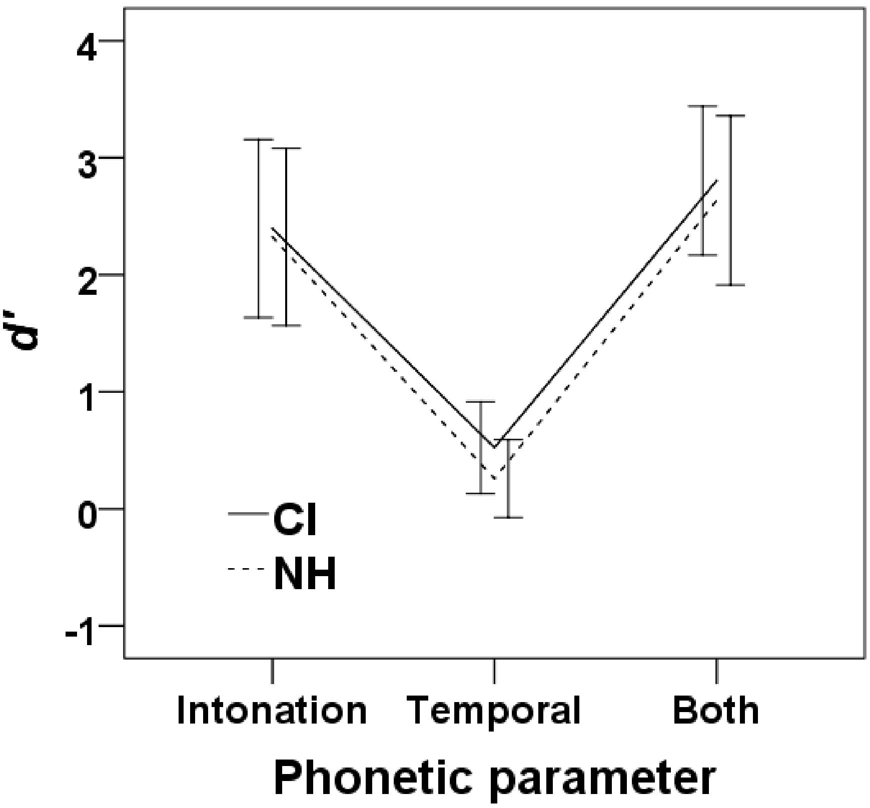


**Figure 2.** Mean *d'* scores split by Phonetic parameter and by participant group (CI or NH) in the Emotion perception test. Participants judged if prerecorded utterances were pronounced with a happy or sad emotion. Phonetic parameters indicate which type of phonetic information was available in the stimulus.

general the groups are almost equally good at distinguishing them. *Main test 3, Focus production.* In the Focus production test, of all trials, 10.9% were missing (missing response or technical error). Table 5 and Figure 4 show mean percentages correct of the three focus positions tested evaluated by a panel of listeners in both participant groups (CI and NH). The mean percentage correct for the CI group was 58.1% and for the NH group 60.4%. A main effect of Focus was 58.1% and for the NH group 60.4%. A main effect of Focus position was found ($F(2, 2774) = 57.00$, $p < .001$, $r = .14$), but not of Group

**Table 4.** Mean percentages correct and standard deviations (in parentheses) per focus position and per participant group (CI or NH) of focus position conveyed in dummy phrases in the Focus production test.

| Emotion | Accuracy mean (SD) | | |
|---|---|---|---|
| | CI | NH | **Total** |
| happy | 70.4 (46.7) | 79.2 (40.7) | **74.6 (43.6)** |
| angry | 70.6 (46.1) | 53.2 (50.0) | **62.1 (48.6)** |
| sad | 62.3 (49.6) | 61.7 (48.7) | **62.0 (48.6)** |
| scared | 46.4 (50.0) | 31.6 (46.6) | **40 (49.0)** |
| **Total** | **62.3 (48.5)** | **57.8 (49.4)** | **60.3 (48.9)** |

($F(1,2774) = 1.94$, $p = .026$) nor an interaction between Focus position and Group ($F(2,2774) = .94$, $p = .39$). These results indicate that the two groups were equally effective at distinguishing the focus positions in their output and that they most likely produced them with similar strategies, given that they were similarly judged by the panel of listeners.

*Correlations among tests and between age and test scores.* Two-tailed correlations between six scores of Non-verbal emotion understanding test and the scores of the Non-word repetition, Emotion perception, Emotion production, and Focus production tests were tested per Group. The six scores of the Non-verbal emotion understanding test were (1) total scores (in numbers of correct responses) for the Face discrimination task (i.e., averaged scores over all four test components) and (2) the Face identification task (total number of items correct over all trials) as well as (3 through 6) percentage correct scores for verbal and pointed responses according to strict and tolerant policies. These subscores of components within Non-verbal emotion understanding were not tested for correlations

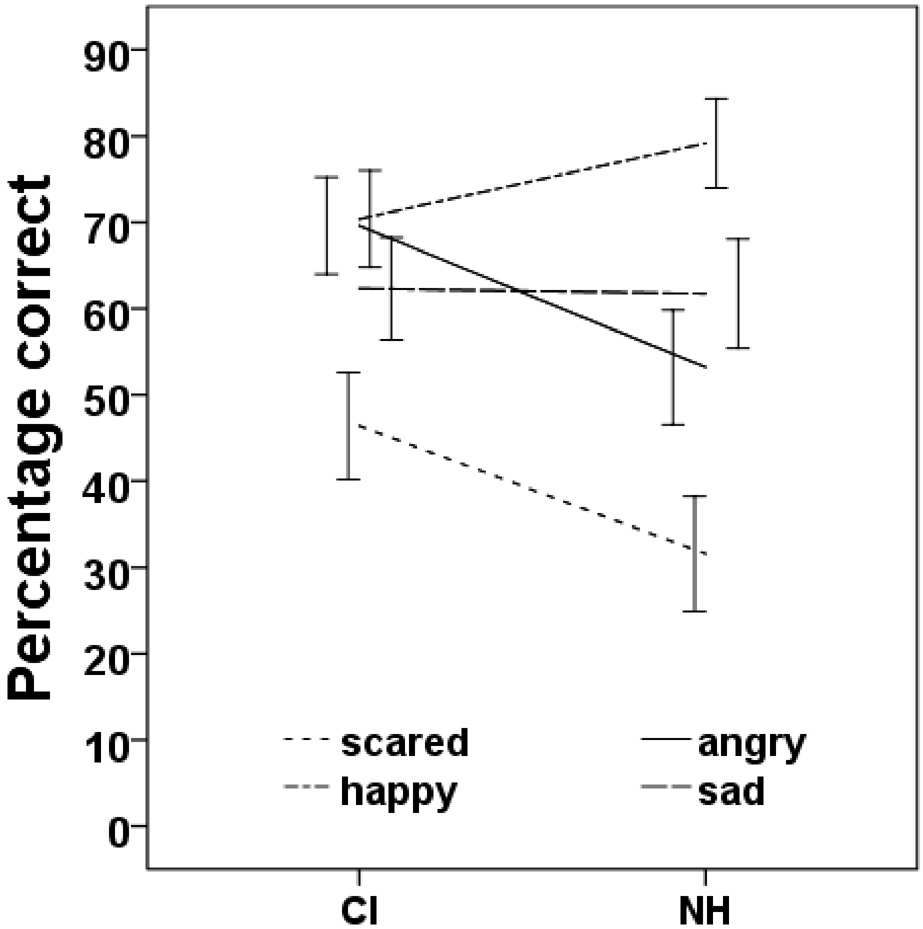


**Figure 3.** Mean percentages correct per emotion and per participant group (CI or NH) of emotions conveyed in dummy phrases in the Emotion production test. Percentages correct were computed by averaging judgements of emotions perceived by a panel of ten naïve adult Dutch listeners with normal hearing.

among each other nor with the Non-word repetition test, but only for correlations with the main tests. Based on Kolmogorov-Smirnov tests and visual inspection of Q-Q plots, we assumed that the distributions per group for Non-word repetition, Emotion perception, Emotion production, and Focus production were largely normal and that distribution for the other scores were not normal. It should be noted

**Table 5.** Mean percentages correct and standard deviations (in parentheses) per focus position and per participant group (CI or NH) of focus position conveyed in dummy phrases in the Focus production test.

| Emotion | Accuracy mean (SD) | | |
|---|---|---|---|
| | CI | NH | **Total** |
| adjective | 70 (45.9) | 72.1 (44.9) | **71.0 (45.4)** |
| noun | 59.6 (49.1) | 59.3 (49.2) | **59.4 (49.1)** |
| both | 44.2 (49.7) | 50.1 (50) | **47.9 (49.9)** |
| **Total** | **58.1 (49.4)** | **60.4 (48.9)** | **59.2 (49.2)** |

that for the CI group the distribution of scores on the Non-word repetition test was marginally significant ($p$ = .063).

Only the following correlations were significant. For the CI group, scores of the Face discrimination were marginally significantly and weakly correlated with Emotion perception ($r$ = .387, $p$ = .046), those of the Face identification task were moderately correlated with Emotion production ($r$ .52, $p$ = .015), strictly judged verbal responses on the Expression task were moderately or weakly correlated with Emotion perception ($r$ = .453, $p$ = .025) and, marginally significantly, Focus production ($r$ = .308, $p$ = .089), respectively, and strictly judged pointed responses were weakly correlated with Emotion production ($r$ = .398, $p$ = .039). In the NH group, strictly judged pointed responses were weakly to moderately and marginally significantly correlated to Focus production ($r$ = .423, $p$ = .054). The correlation between Emotion perception and Emotion production was marginally significant for the CI group ($r$ = .523, $p$ = .067) whereas it was not for the NH group ($r$ = −1.44, $p$ = .656), and the correlation between Non-word repetition and Emotion production was marginally significant for NH group ($r$ = .543, $p$ = .068) whereas it was not for the CI group ($r$ = .017, $p$ = .96).

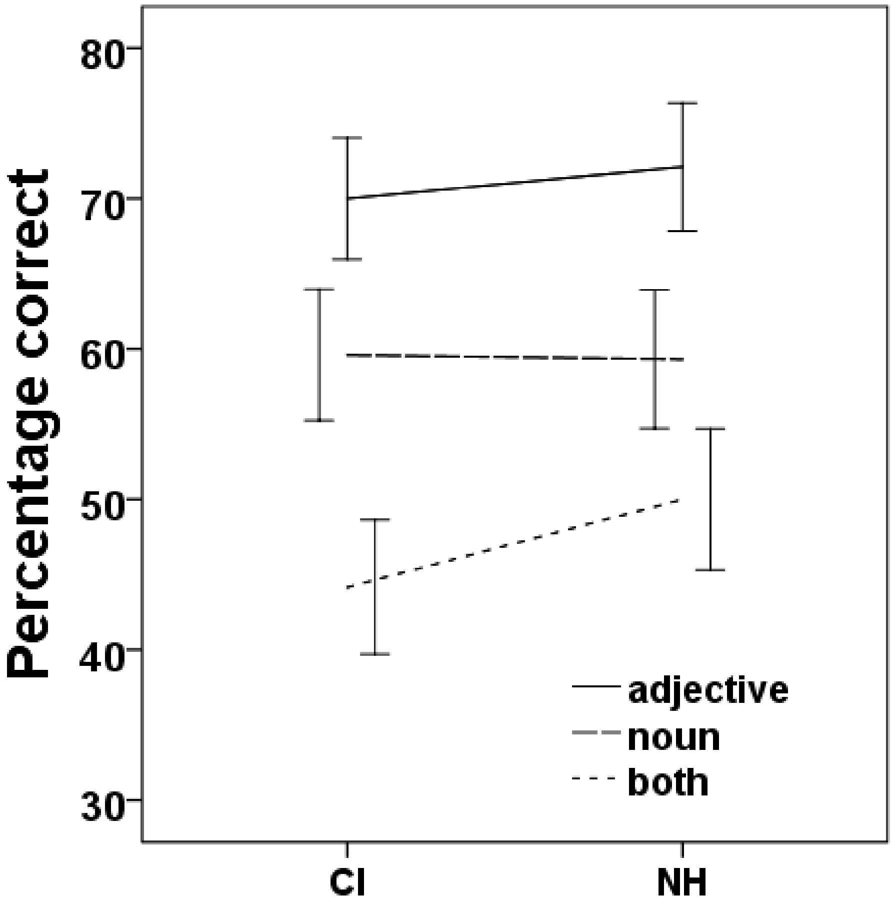


**Figure 4.** Mean percentages correct per focus position and per participant group (CI or NH) of focus positions conveyed in dummy phrases in the Focus production test. Percentages correct were computed by averaging judgements of emotions perceived by a panel of ten naïve adult Dutch listeners with normal hearing.

Finally, correlations were run between main test scores on the one hand, and activation age, hearing and chronological age, on the other hand. In the CI group, the only significant correlation was between hearing age and Emotion production ($r$ = .028, $p$ = .542). In the NH group (where chronological age is by definition equivalent to hearing age), there was a marginally significant correlation between age and Emotion production ($r = .470$, $p = .061$).

These results show that in general, the capacities tested in the different main tasks seem unrelated to each other but that there is a trend towards emotion production being predicted by emotion perception skills for CI (but not NH) children and by Non-word repetition skills for the NH (but not the CI) group. Moreover, partly as a trend, scores on the Emotion perception and production and Focus production are to a limited degree predicted by some non-verbal emotion understanding scores, although more so for the CI than for the NH group. Age measures were to some extent correlated with Emotion production but not with other main test scores.

## 6.4 Discussion

The aim of this study was to test the capabilities and cue-weighting strategies of a small group of children with cochlear implants and a control group of normally hearing children, matched for hearing age, on both the perception and production of (emotional and linguistic) prosody, controlling for the level of non-verbal emotional understanding and general linguistic capacities, and to test for correlations between scores of main tests and between baseline and main tests. To our knowledge, this study was the first to test perception and production of emotional (and linguistic) prosody in the same cohort of pediatric CI recipients. Moreover, effectiveness of the non-imitative production of emotions was never tested in this population. Although, contrary to our hypotheses, the two groups performed generally in a similar way, some differences were observed that coincided with our expectations.

Our first set of hypotheses was (A1) that prosody perception and production scores within participants were correlated (A2) that that effect was larger within than across prosody type (linguistic vs. emotional) and (A3) that that effect was larger for linguistic than for emotional prosody. Hypothesis A1 was confirmed to a limited degree. In the CI group, but not the NH group, Emotion perception performance moderately and marginally significantly predicted Emotion production performance. Other correlations, however, were either very weak and/or not significant. This result is in support of hypothesis A2, since the only correlation of any significance involves within-prosody type (emotional linguistic) and between-prosody correlations were not found. As Focus perception, however, could not be analyzed, it remains unknown if this holds for linguistic prosody as well. For the same reason, Hypothesis A3 cannot be confirmed nor rejected.

The trend for a link between emotion perception and production in the CI group supports results by Nakata et al. (2012) who found a correlation for five- to thirteen-year-old recipients between imitative emotion perception and production scores. The trend, if reflecting an actual effect, provides some support for the view that in this population better prosody perception skills allow better prosody production skills (e.g., Nakata et al., 2012), at least within the domain of emotion. This would entail that the production of emotions cannot develop and function entirely independently from their perception, whereby the independence stance would stem from the idea that the way to distinguish vocal emotions in production is not (sociolinguistically) acquired but innate (Scherer, Banse & Wallbott, 2001). Instead, thus, our results argue in favor of the opposite hypothesis, stating that vocal expression of emotions is at least partly learned. This is also consistent with the fact that in the present data for the NH children, no such trend was observed, as, naturally, they have received normal input since birth and their variation in skills of emotion perception and production distinction might be due to other factors, such as personality factors, instead of perceptual acuity. It has to be kept in mind, however, that due to the small sample size, personality factors, for instance, might have played a role in the experimental group as well, for instance representing a wider variation than for the NH group.

As our second set of hypotheses, we expected (B1) that scores per prosody type (linguistic or emotional) were influenced by their respective general scores for linguistic and emotional capacities, (B2) but that this effect was larger for linguistic than for emotional prosody. As for prediction by the linguistic baseline test (Non-word repetition), a marginal correlation between Non-word repetition and Emotion production was found in the NH group, but no other correlations. As for the emotional baseline test (Non-verbal emotion understanding), in the CI group, Emotion perception and Emotion production each correlated with scores on two of the Non-verbal emotion understanding subtests, namely Face discrimination and strictly judged verbal responses for the Expression task, on the one hand, and Face identification and strictly judged pointed responses in the Expression task, on the other hand, respectively. In the NH group, there was only a marginal correlation between Focus production and strictly judged pointed responses in the Expression task. Therefore, Hypotheses B1 (concerning emotion tasks) and B2 are partly confirmed for the CI group, but not for the NH group.

These results complement and in part contradict those by Wiefferink et al. (2013), from whose battery of non-verbal emotion understanding tests those in the present study were adopted. Whereas Wiefferink et al. (2013) reported a delay in emotional development in CI children, in the present study, no difference was found. This lack of a difference is, however, consonant with the normal emotional development found by Mancini et al. (2016). As in the study by Mancini and colleagues, whose implanted participants were between 4 and 11 years old, it is highly probable that this is due to the fact that the current study tested older children (between 6 and 12 years old). Although these older children did not reach ceiling level performance on all subtests, the tests might be less sensitive to possibly more nuanced differences in emotional capacities at these ages. Nevertheless, importantly, the similarity in performance of the two groups on non-verbal emotional understanding tests suggests that CI children's emotional capacities (partly) are at a level comparable with that of their peers near – at the latest – the end of primary school.

A difference with both the studies by Wiefferink et al. (2013) and Mancini et al. (2016) is that for CI children no correlation between performance on verbal emotion tasks and general language level was found. A connection between those faculties was explained by Flom and Bahrick (2007), cited in Mancini et al. (2016), by assuming that language helps naming emotions and linking them to external referents (objects and events) and that the temporal synchrony between vocally (i.e., prosodically) and non-vocally expressed emotions (e.g., faces) and external referents is required for leaning to distinguish emotions. Given this hypothesis, the present lack of a correlation between general linguistic and verbal emotion tasks might be accounted for by assuming that the linguistic experience of the CI children, whose linguistic level was less advanced than that of the control group but as a small effect, was sufficient for talking about emotions, linking them to external referents and learning the synchrony with facial expressions. It has to be noted that CI and NH participants were matched for hearing age (not chronological age) and the experimental group therefore had more chance to gain experience than the control group with learning to distinguish and express – verbal and non-verbal – emotions.

Our third set of hypotheses was (C1) that CI activation age would negatively correlate with prosody capacities, but (C2) that this effect would be larger for linguistic than for emotional prosody. In the CI group, CI activation age nor chronological age was found to predict outcomes, but hearing age did moderately correlate with Emotion production. In the NH group, age was marginally significantly and weakly to moderately correlated with Emotion production. The hypotheses are therefore not confirmed because any effect observed is related to emotional and not linguistic prosody processing. Increasing experience with the implant (hearing age) did improve emotion perception, but hearing age negatively correlated with activation age, obscuring conclusions about the separate effect of either factor. Samples were too small to perform partial correlations.

These results suggest that increasing time in sound and possibly an earlier onset of stable hearing (in this study defined as birth for the NH group and age at activation for the CI group) help improve emotion production performance. The fact that emotion perception was not found to be predicted by this factor would suggest that emotion perception capacities mature independently of hearing experience, whereas emotion production capacities do develop as a function of it. The finding that emotion production is at most indirectly dependent on other capacities resonates with results from the study by Bergeson and Chin (2008), whose six- to ten-year-old implanted children's emotion prosody imitation performance showed no correlations with intelligence scores and general linguistic level. This at first glance seems inconsistent with our other result that emotion production skills are correlated with emotion perception skills; however, the two lines of results could be reconciled by the assumption that emotion perception capacities are a necessary but not a sufficient requirement for emotion production capacities, whereby one of the other possible requirements are sufficient emotional capacities (as also observed in the present study).

Our final set of hypotheses was (D1) that for the perception of prosody, CI participants would rely more heavily on temporal cues as opposed to F0 cues than NH participants do, but (D2) that this effect would be stronger for linguistic than for emotional prosody. In as far as this study tested these hypotheses, they were not confirmed. The recipients weighted their cues in the same way as the children from the control group, namely by relying almost entirely on F0 cues, disconfirming Hypothesis D1. Hypothesis D2 was not tested because the Focus perception test did not yield analyzable results. That hypothesis therefore remains to be investigated in future research.

The similarity in emotion perception performance and cue weighting strategy between CI and NH children is in marked contrast with earlier research, where children of different ages showed poorer emotion perception performance (Geers et al., 2013; Luo et al., 2007; Nakata et al., 2012) and a heavier reliance on temporal vs. F0 information by CI users (Meister et al., 2011; O'Halpin, 2009). It might be the case that the happy and sad stimuli used happened to have relatively pronounced differences in intonation contour and/or register, allowing even CI users, who have poor F0 resolution, to reach ceiling level when only F0 information was present, possibly also diverting their attention from temporal information when only temporal information was present (in the Duration condition). More difficult tasks, with less exaggerated renderings of emotions and/or with more emotions, might bring to light more subtle differences in cue weighting strategies between these groups. Nevertheless, it is remarkable that in this study the F0 information was sufficient for CI children to distinguish emotions at a level equal to that of their NH peers.

### ***Shortcomings and suggestions for future research***

A number of shortcomings of this study have to be taken into account that warrant prudence in interpreting the data. First of all, the sample size, with two groups of thirteen participants, was small as a result of the limited availability of implanted children passing the inclusion criteria, compromising generalizability to other pediatric recipients.

Second, possibly, the cohort has self-selected for the better-performing children because parents who feared their children might perform sub-optimally might for that reason not have responded to the invitation. This may have contributed to the fact that the implanted children performed within the norm on several (sub)tests. The fact that effects and tendencies have been found in the present sample suggests that stronger effects could be found in larger studies.

For the above reasons, research using larger sample sizes with implanted children with a broad range of linguistic and maturational developmental levels (i.e., chronological and hearing ages) is likely to extend and strengthen the present results. Further, a study in which early and late implanted children, compared to three groups matched for hearing age and chronological age to both experimental groups would further complement the current study by separately testing the roles of duration of CI experience and general maturational development for emotional and linguistic language skills.

Third, a limited set of stimuli was used in all tests so that the results of the tests would allow a comparison. Moreover, the stimuli's variants with different cue availability (intonation, temporal information or both) were highly controlled in order to test the role of the respective types of phonetic information irrespective of the contents of the stimuli. The sole speaker recruited to record the stimuli, however, will have idiosyncratic prosodic characteristics (Kraayeveld, 1997); the production by another speaker or of other stimuli might have brought about different weightings of temporal and intonation information. It can therefore not be excluded that the cue reliance mechanism found in this study is specific to the stimuli used.

The baseline test for non-verbal emotion understanding used in this study might not have been sensitive nuances in emotional development that could influence performance on linguistic prosody tests. More challenging tests, such as involving higher emotional skills such as beliefs and moral values, in combination with the tests of the perception and production of irony, surprise and deception and acoustic measurements of elicited utterances, might capture fine-grained differences in emotional verbal development in this population.

### *Conclusions*

In this research, we tested the perception and production of emotional and linguistic prosody by six- to twelve-year-old children with cochlear implants and normally hearing, hearing age matched children. It has to be noted that linguistic prosody perception (focus perception) could not be analyzed. The following conclusions resulted from the study.

1) Emotional prosody perception and production scores were weakly and moderately significantly correlated for CI children but uncorrelated for NH children, suggesting that higher perception skills allow higher production skills and that emotion production is partly learned (as opposed to innate).
2) For CI children but not NH children, emotion perception and production scores were predicted by non-verbal emotional understanding performance. No such correlation was found for linguistic prosody. For NH children, only marginal contra-modal (from emotion to focus and vice versa) correlations were found. Our data showed no overall performance level difference between groups, suggesting that these children either never experienced deviations for the tested capacities or if they had had any delay, that has become irrelevant by the age at testing.
3) Hearing age (itself correlated with activation age for the CI children) weakly predicted emotion production, but not emotion perception, performance in both groups, suggesting that increasing time in sound has a favorable effect on vocal emotional expression, possibly requiring independently maturing emotion perception skills.
4) For emotion perception, CI and NH children adopt the same cue-weighting strategies, relying almost entirely on F0 information as opposed to temporal information, and perform at the same level of accuracy.

**Acknowledgments**

We are grateful to Jos Pacilly, engineer at the Phonetics Laboratory at Leiden University, for his technical support at many stages of this research. We are also grateful to Walter Verlaan, engineer at Leiden University Medical Center, for his assistance in developing the experimental setup. We also thank all participants for their willingness to cooperate.