

ARTICLE

# Prosody perception and production by children with cochlear implants

Daan J. VAN DE VELDE<sup>1,2\*</sup>, Niels O. SCHILLER<sup>1,2</sup>, Claartje C. LEVELT<sup>1,2</sup>,  
Vincent J. VAN HEUVEN<sup>3</sup>, Mieke BEERS<sup>4</sup>, Jeroen J. BRIAIRE<sup>4</sup>, and Johan H. M. FRIJNS<sup>2,4</sup>

<sup>1</sup>Leiden University Centre for Linguistics, Leiden University, Van Wijkplaats 3, 2311 BX, Leiden, <sup>2</sup>Leiden Institute for Brain and Cognition, Postbus 9600, 2300 RC, Leiden, <sup>3</sup>Department of Hungarian and Applied Linguistics, Pannon Egyetem, 10 Egyetem Ut., 8200 Veszprém, Hungary, and <sup>4</sup>Leiden University Medical Center, ENT Department, Postbus 9600, 2300 RC, Leiden

\*Corresponding author: Daan van de Velde, Leiden University Centre for Linguistics, Leiden University, Cleveringaplaats 1, 2311 BD, Leiden, the Netherlands. E-mail: [dj.van.de.velde@hum.leidenuniv.nl](mailto:dj.van.de.velde@hum.leidenuniv.nl)

(Received 14 April 2017; revised 23 March 2018; accepted 19 August 2018;  
first published online 18 October 2018)

## Abstract

The perception and production of emotional and linguistic (focus) prosody were compared in children with cochlear implants (CI) and normally hearing (NH) peers. Thirteen CI and thirteen hearing-age-matched school-aged NH children were tested, as baseline, on non-verbal emotion understanding, non-word repetition, and stimulus identification and naming. Main tests were verbal emotion discrimination, verbal focus position discrimination, acted emotion production, and focus production. Productions were evaluated by NH adult Dutch listeners. All scores between groups were comparable, except a lower score for the CI group for non-word repetition. Emotional prosody perception and production scores correlated weakly for CI children but were uncorrelated for NH children. In general, hearing age weakly predicted emotion production but not perception. Non-verbal emotional (but not linguistic) understanding predicted CI children's (but not controls') emotion perception and production. In conclusion, increasing time in sound might facilitate vocal emotional expression, possibly requiring independently maturing emotion perception skills.

**Keywords:** cochlear implants; prosody; phonetic cues

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 message-related (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

© The Author(s), 2018. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

hemisphere and linguistic prosody with both hemispheres (Witteman, van IJzendoorn, van de Velde, van Heuven, & Schiller, 2011). Second, they are phonetically different. 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. More generally, Gussenhoven (2004, pp. 49–70), referring to design features of language by Hockett (1960), argued that non-linguistic intonation lacks arbitrariness (i.e., is iconic), discreteness, and duality of patterning, whereas linguistic intonation does have these features. From a production perspective, the phonetic difference is that linguistic prosody is discrete, whereas emotional prosody is gradient (Gussenhoven, 2004, pp. 49–70), and that emotional (Scherer, Banse, Wallbott, & Goldbeck, 1991) and linguistic information, such as marking of new vs. old information in sentences (Chen, Den Os, & De Ruiter, 2007) and sentence type (Kord, Shahbodaghi, Khodami, Nourbakhsh, & Jalaei, 2013), may be conveyed by specific prosodic structures. From a perception perspective, the two prosody types have different relative importance of F0 vs. duration parameters (Murray & Arnott, 1993; Sityaev & House, 2003; Williams & Stevens, 1972). The greater intuitiveness of emotional prosody is supported by the observation that its vocal expression is cross-linguistic (Scherer, Banse, & Wallbott, 2001; however, universality is controversial; Scherer *et al.*, 1991).

In normal hearing (NH) children, prosody plays a crucial role in language acquisition. Due to child-directed speech, infants focus on important linguistic and paralinguistic aspects of speech (Liu, Kuhl, & Tsao, 2003). Furthermore, prosody might initiate speech segmentation, thus helping to trigger word learning (Johnson & Jusczyk, 2001). Children use prosody to infer, or (in production) show, emotional states as early as the first year of life (Flom & Bahrick, 2007; Scheiner, Hammerschmidt, Jürgens, & Zwirner, 2006), but the development of their mastery in some studies extends well into primary school (Aguert, Laval, Lacroix, Gil, & Le Bigot, 2013; Peppé, McCann, Gibbon, O'Hare, & Rutherford, 2007). For certain forms of linguistic prosody, this acquisition seems to develop over a longer time, although with differentiation between languages and prosody types (Chen, 2007; de Ruiter, 2010; Filipe, Peppé, Frota, & Vicente, 2017; Ito, Bibyk, Wagner, & Speer, 2014). Linguistic intonation acquisition has been found to correlate with more general measures of linguistic development (Wells, Peppe, & Goulandris, 2004). Moreover, clinical populations such as cochlear implant (CI) users and children with autism or aphasia show prosodic deficits (Baltaxe, Simmons, & Zee, 1984). Children with CIs, however, have been reported to be able to recognize emotions in faces despite having difficulty recognizing vocal affect (Hopyan-Misakyan, Gordon, Dennis, & Papsin, 2009).

The current study compares linguistic and emotional prosody perception and production capacities of children with CIs with normally hearing children. Inherent group and prosody type differences suggest correlations within prosody type and group (hypothesis group A below). The above observations about the development of prosody are relevant for this study in the sense that emotional prosody acquisition might be less directly linked to more general development than linguistic prosody acquisition is. This is tested by finding correlations between prosodic and more general developmental tests (hypothesis groups B and C and hypothesis A3).

### *Language in children with cochlear implants*

Children with cochlear implants experience delays or deviations in their oral (productive and perceptual) linguistic and socio-emotional development relative to

normally hearing peers (Geers, Nicholas, Tobey, & Davidson, 2016; Geers, Tobey, Moog, & Brenner, 2008; Robinson, 1998; Wiefferink, Rieffe, Ketelaar, De Raeve, & Frijns, 2013). 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; however, see Iverson, Smith & Evans, 2006; Välimaa, Määttä, Löppönen, & Sorri, M. J., 2002; Välimaa, Sorri, Laitakari, Sivonen, & Muhli, 2011), 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 *et al.*, 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 *et al.*, 2014; however, see Meister *et al.*, 2009). Problems with the production of speech have also been observed, including voice quality (Ubrig *et al.*, 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).

The importance of speech perception and production for CI children's linguistic development has been demonstrated in different lines of research. First of all, in a series of studies testing 181 implanted children, speech perception and production performance explained 42% of overall total language scores and as much as 63% when split for overall spoken language scores (Geers, Nicholas, & Sedey, 2003). Furthermore, problems in those areas have been associated with delays in socio-emotional development. Wiefferink *et al.* (2013) tested Dutch CI and NH two-and-a-half- 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 some aspects of linguistic development can predict aspects of emotional development. Mancini *et al.* (2016), however, found that 79% of their cohort of 72 CI children, aged four to eleven years, showed normal emotion understanding skills. The differences with Wiefferink *et al.*'s (2013) results were attributed to discrepancies between the participant groups: Mancini *et al.*'s (2016) 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 on 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 the emotional and linguistic development of CI children. Collectively, this literature warrants more research tracking the development of speech perception and production in CI children.

*Prosody perception in CI users*

For pediatric as well as adult CI users, several aspects of the perception and production of emotional and linguistic prosody have proven relatively problematic compared to comprehension of sentences (e.g., Helms *et al.*, 1997).

First of all, 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 *et al.* (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.

As for the perception of linguistic prosody, Meister *et al.* (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 the 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).

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 challenges. Meister, Landwehr, Pyschny, Wagner, and Walger (2011) measured DIFFERENCE LIMENS (DL, i.e., smallest discernable differences) 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 so in the dynamic and probably even less so in the temporal domain, and that this has repercussions for the type of prosodic information that they adequately receive. In this study, this resolution hierarchy is tested in CI children and compared to NH children (hypothesis group D).

### *Prosody production in CI users*

A small number of studies have addressed the issue of prosody production by CI users. Lyxell *et al.* (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 five to thirteen 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, statements and questions 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 respects. There is, however, mixed evidence regarding the question whether good perception skills are required for good production skills.

### *Current study*

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 comparing them to a control group of NH peers. The development of linguistic and emotional prosody has never been clearly contrasted. This line of research needs to be undertaken because the perceptual 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 socioeconomic status was also gathered. We tested the following hypotheses.

- (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) Due to inherent differences between the prosody types (linguistic vs. emotional), that effect is larger within than across the prosody type, and (A3) larger for linguistic than for emotional prosody because emotional production, due to 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 development, (B2) but this effect is larger for linguistic than for emotional prosody, since components of linguistic development are supposedly to a relatively high degree acquired as an integrated system of rules.
- (C1) Assuming a possible effect of more general maturation on linguistic, including prosodic, skills (hypothesis group B), CI activation age negatively correlates with prosody development, 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 this effect to be stronger for linguistic than for emotional prosody.

In summary, we investigated whether 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 development, and if CI users used different cues for prosody perception and production than the NH control group.

## Method

The test battery comprised several components for every child: a quartet of main tests, a familiarization test for that quartet, four baseline tests, and a parent questionnaire.

The quartet of main tests comprised tests on emotion perception, focus perception, emotion production, and focus production, the order of which was randomized across participants. This block was preceded by one round of familiarization with the names of the stimuli (colours and objects). The additional four baseline tests had 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 capacities at the moment allowed. The non-verbal emotion understanding comprised two tests from a battery designed to assess social-emotional development in normally hearing children 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 development, which might or might not correlate with scores on tests gauging prosody development. Finally, the parents or caretakers were asked to complete a questionnaire about their socioeconomic 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).

### **Participants**

Thirteen implanted children and thirteen children with normal hearing participated in this study. They were matched on gender (eleven boys in both groups) 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 a 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. Of the CI children, six sometimes used hearing aids before implantation, one regularly, and for six of them it was unknown. 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](#).

Results of the parent questionnaire were as follows. 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 of finished education (an 8-point scale ranging from elementary school to university level education) and their income category (a 5-point scale ranging from 'lower than €15,000' to 'higher than €60,000' per year), 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 SNL teachers. Dutch Sign Language

**Table 1.** Demographic and Implant Characteristics of Recipients.

Subject number (gender)	Chronological age	Estimated age at hearing loss onset	Estimated duration of deafness (months)	Age at first CI activation	Hearing age (months)	Aetiology	Implanted ear(s)	Current implant type	Current speech processor
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 McCullough	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

Notes. 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.

(Nederlandse Gebarentaal or NGT; Sign Language of the Netherlands or SLN) is the sign language used by deaf people in the Netherlands. Three parents indicated that the acquisition of spoken Dutch was simultaneous with that of SNL and two parents had not reported the acquisition onset age for SNL. 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 for 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.

### 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 (author 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 and Focus perception test, all trials were based on six object and four colour names in Dutch: *auto* 'car', *bal* 'ball', *ballon* 'balloon', *bloem* 'flower', *schoen* 'shoe', *stoel* 'chair', *blauw* 'blue', *geel* 'yellow', *groen* 'green', 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 colour bias, to avoid anomalies such as green bananas and blue trees; (4) nouns had common gender, so they had the same article and adjectival suffix; and (5) the nouns were known by at least 86% of children aged 2;3 as tested by a questionnaire with 961 (pairs of) parents and listed in the Lexilijst (Schlichting & Lutje Spelberg, 2002). According to that questionnaire, the colours were known by between 47% and 63% of children of that age. However, we chose these items because they were the four most frequent colours known by young children, because our participants had a higher hearing age than 2;3, and because they were familiarized with the stimuli before the test phase. Despite criterion (1) above, some presence of voiceless segments in the item list was unavoidable, and priority was given to the criteria of familiarity and natural colour-neutrality. 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 had an advantage.

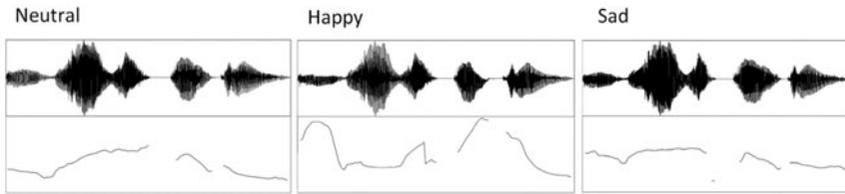
In the Emotion perception test, all 24 combinations of the six objects and four colours were produced in a happy and a sad variant. The phrases followed the template *een* [colour] [N], where *een* is the singular indefinite article. They were between 1.38 and 1.93 seconds long, with an average duration of 1.72 seconds for happy and 1.62 seconds for sad phrases. In the Focus perception test, there were 12 colour-object combinations, half of which carried focus on the colour adjective and the other half of which had focus on the noun. The phrases followed the template *een* [colour] [N] *en een* and were between 1.46 and 1.74 seconds long, with a mean

duration of 1.56 seconds for colour focused phrases and 1.58 seconds for object focused phrases. The trailing phrases *en een* ‘and a’ were intended to prevent phrase-final prosody on the noun and were pronounced with a question intonation in order to elicit a response. It has been reported elsewhere (van de Velde *et al.*, 2017) that the emotions and focus positions, taking 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. That reference also provides acoustic information about the stimuli in the various conditions of both tests. Figure 1 shows an example of the waveform and intonation contour of a neutral, happy, and sad variants of stimuli in the Both condition.

Sentences in both tests were all manipulated into three extra variants by CROSS-SPLICING aspects of the prosody from the non-neutral (emotional or focused) stimuli to the same neutral equivalents (the Cue condition): (1) only the F0 contour (F0 condition); (2) only the durations of the segments (Duration condition); and (3) both the F0 contour and the segment durations (Both condition). This was done in order to control the cues available to the participants. Because single neutral variants (i.e., one single variant for the two emotions or focus positions) constituted the bases of the stimuli, judgements by participants could only be based on F0, segment durations, or both, respectively. Except for these cues, the two emotions or focus positions were identical, since the underlying segmental material was identical for both emotion or focus position variants of a given phrase. For the Focus production test, stimuli of the format *Is dit een* [colour] [N]? ‘Is this a [colour] [N]?’ were recorded. The speaker was asked to prevent expressing focus on one of the content words specifically. As focusless sentences cannot be produced, this strategy in reality resulted in double-focused questions.

In all relevant tests, response options were represented with additional images. Pictures recurring in different tests were those depicting the auditory noun and colour 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 colours using Microsoft Paint in order to be able to contrast coloured objects with each other. All children were familiarized with the visual stimuli before testing by showing all colour 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 (the pre-main-tests familiarization test). 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 colour/object combinations. The auditory stimuli were always the identical tokens of the same phrase and the visual materials were 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 the Focus discrimination test, the response options were a coloured square (for the colour response) and a black and white line drawing of an object (for the noun response). In baseline test 4, the Non-word repetition test, stimuli consisted of nonsense words in a carrier phrase presented as the supposed words for fantasy toys of which



**Figure 1.** Example of the waveform and intonation contour (scaled between 75 and 500 Hz) of stimuli in the Both condition of the Emotion perception test, produced using *Praat*. Shown are neutral, happy, and sad variants of *Een blauwe auto* 'A blue car'. Total stimulus durations (1.93 s. for neutral, 2.01 s. for happy, and 2.24 s. for sad) as well as allophone durations were different between emotion conditions.

coloured 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, 2016). The nonsense words were four stimuli of each 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.

The criteria for the phonological composition of the nonsense words (based on Dollaghan & Campbell, 1998), were as follows: (1) non-final syllables were CV; final syllables were CVC. (2) To ensure that non-word repetition would not be affected by a participant's vocabulary knowledge, the constituent syllables (CV or CVC) preferably did not correspond to a Dutch word. Of the 60 syllables in the set, 19 were not non-words, but all but two of those had a frequency of 2 or less per million according to the Celex database (<http://celex.mpi.nl/>). (3) Syllables only contained phonemes that even atypically developing children with a chronological age of 2;8 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). (4) Syllables contained only tense vowels /i, u, e, o, a/ (i.e., the unmarked subset of five vowels from the vowel inventory of Dutch; lax vowels were excluded since these cannot occur at the end of syllables so that a following onset C must either be geminated or doubly associated (e.g., van Heuven, 1992, and references therein). (5) To limit syllabic positional predictability, consonants, except /s/, occupied only positions (onset vs. coda) in which they occurred less than 32% of their occurrences in realized (not underlying) forms according to token frequencies retrieved from the Corpus Gesproken Nederlands (<http://lands.let.ru.nl/cgn/>) (Van Oostendorp, p.c.). (6) For independent recall, any C appeared only once in a word, primary stress was realized on the final syllable, and secondary stress on the initial syllable, except for trisyllabic nonwords, where it was the other way round.

Practice stimuli were different from the experimental stimuli. The carrier phrase of all nonwords 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 nonwords can be found in the 'Appendix'.

### 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 required it. 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 Stimulus 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 the 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 tests except 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 the 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 colour.

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 dB SPL 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 of a large portion of the combined stimuli (thus from the same speaker) of the experiments. Note that the use of headphones was not an option as they would interfere with the 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 (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 fearful) and was asked to tell how the protagonist, always shown from behind the head to avoid 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 emotions 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/colour combinations by selecting a picture on the screen. The target position was counterbalanced, as were the position and type of the distractors (only different colour, only different object, both different). 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 colour blindness, only two colour 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 colour/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*

The 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 picture illustration of the pronounced phrase (e.g., a picture of a blue ball for the pronounced phrase *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: Focus perception*

Participants heard a phrase pronounced with the focus either on the colour (e.g., ‘a BLUE ball and a’) or the noun (‘a blue BALL and a’) and indicated which word they felt carried focus by selecting a coloured square or an uncoloured object. They were trained for this procedure by means of live examples and cards, whereby the experimenter exaggerated emphasis on the adjective or the noun and invited the child to indicate if the coloured square or the uncoloured line drawing of the object belonged to that. The child was not corrected, as different children might perceive emphasis differently. An extra trial, up to a maximum of three, was performed, however, when the child gave the impression of not understanding the task. There were two counterbalanced blocks of 24 randomized trials in four conditions of six phrases. The four conditions were the three Cue conditions plus a condition with no manipulation at all. The experimental part was preceded by four warm-up trials, one in each condition. Trials had a structure similar to those in the Emotion perception test. Response options were not swapped because that would have created a conflict with the order of the words in the stimuli.

*Main test 3: 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 were eight trials, namely two objects to be named with each of the emotions ‘happy’, ‘sad’, ‘angry’, and ‘fearful’. 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 [colour] [N]?’ either matched (half of the stimuli) a picture they produced or contrasted in the colour 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 set-up to those of the Emotion and Focus perception tests.

*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. The alpha decision level was  $p = 0.05$  and a  $p$  value of .05 or lower was assumed to be significant. Analyses were performed using SPSS version 23.0 (IBM, 2012). Effect sizes are reported only for two-way comparisons and not for less fine-grained (i.e., more global) comparisons, because the latter are not the focus 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, fearful) and a response counted as accurate if and only if that exact emotion was chosen. In the tolerant policy, only a distinction between positive (happy) and negative (angry, sad, fearful) emotions was made. Positive or negative vocabulary other than the expected emotion labels was 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, addition of a phoneme and substitution by another phoneme counting as an error. In case of omitted or added syllables, utterances were aligned with the target in such a way as to minimize the number of errors. Subsequently, the number 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 and 3: Emotion perception test and Focus perception test*

Because, in the Emotion test, only two response options were available, following Signal Detection Theory, scores were transformed into hit rates, 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  $100\%/2N$ , where N is the number of items in the cell (24). Results are presented as  $d'$  scores. Data were subsequently subjected to a mixed-model ANOVA, with Phonetic Parameter as the within-subjects variable and Group as the between-subjects variable. The Focus perception test appeared to be ill understood by a too large number of participants. Moreover, data were not properly recorded due to a technical error. Results for this test will therefore not be discussed. We did, however, consider it useful to not withhold its methodology.

#### *Main tests 2 and 4: Emotion production and Focus production*

Participants' verbal responses in the Emotion and Focus production tests were evaluated by a single panel of ten Dutch adults with a mean age of 27.3 years who did not present a hearing loss of over 40 dB HL 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, fearful) was conveyed independent of the contents of the utterance. In the Focus test, they judged which of three focus positions (colour, object, or both) was accented. Another condition of the Focus production test, in which the question posed to the children corresponded in both colour and object to the image displayed, was not further analyzed. In this test, the procedure of the production task was explained and listeners were asked to imagine which question the speaker's utterance was a response to, so that they would judge the phrasal accents as contrastive focus realizations (which is how they were intended by the speaker). 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. The use of raters was preferred over phonetic or phonological analysis of the realization of the emotions and focus positions, respectively, because the goal was to ascertain to what extent the speakers were able to convey the targets. Phonetic or phonological analysis might have revealed realization differences but that would not have determined if those characteristics were sufficient and appropriate for the intended emotion and focus position. The raters' tasks were designed to minimize 'phonetic' and maximize 'linguistic' listening. This is most clear in the focus rating, where a possible alternative would have been to ask raters to indicate the part of the sentence carrying the main prominence. This would have run the risk, however, of raters focusing on some but not all phonetic features that are relevant for the focal contrast. Since speaker groups might have differed in their realizations but still both be effective, such a rater strategy would miss possibly relevant aspects in one or both groups. For these reasons, a metalinguistic task was administered.

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 datapoints 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).

## Results

Results per test (baseline and main tests) are discussed first. Results of baseline tests should reveal if baseline capacities differ between groups; in the results of main tests, the central issue is the comparison between groups including the possible interaction between Group and condition. Group differences in main tests without group differences in baseline tests are an indication of specific developmental group differences in prosody. After the separate test results, correlations on the level of the individual participant are run between results of different tests in order to find possible explanations for groups' results.

### *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 ( $U = 1230.5$ ,  $z = -1.17$ ,  $p = .24$ ,  $r =$

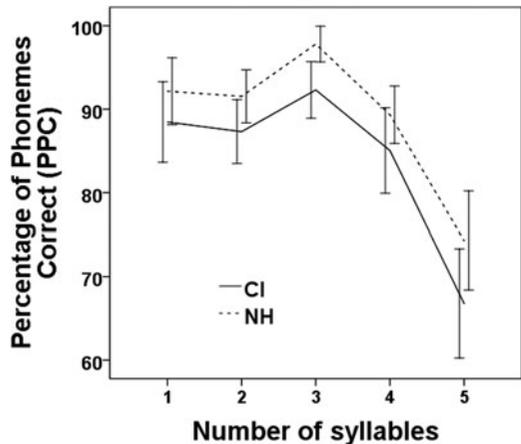
-0.23) or any of the pairs separately according to Mann-Whitney  $U$  tests (cars vs. flowers:  $U = 84.5$ ,  $z = 0$ ;  $p = 1$ ,  $r = 0$ ; faces with glasses vs. hats:  $U = 71.5$ ,  $z = -1.44$ ,  $p = .51$ ,  $r = -0.28$ ; negative vs. positive faces:  $U = 77.0$ ,  $z = -0.56$ ,  $p = .72$ ,  $r = -0.11$ ; angry vs. sad faces:  $U = 74.5$ ,  $z = -0.56$ ,  $p = .61$ ,  $r = -0.11$ ; exact significance). In the Face identification task, no effect of group on the number of correct responses was found either ( $U = 53.0$ ,  $z = -1.8$ ,  $p = .11$ ,  $r = -0.35$ ; exact significance). In the Expression task, no effect of group on mean accuracy scores was found for strict ( $U = 4724.5$ ,  $z = -1.0$ ,  $p = .32$ ,  $r = -0.20$ ) and, although approaching significance, tolerant ( $U = 4892.5$ ,  $z = -1.8$ ,  $p = .074$ ,  $r = -0.35$ ) verbal responses, nor for the strict ( $U = 5267.5$ ,  $z = -0.26$ ,  $p = .79$ ,  $r = -0.051$ ) and tolerant ( $U = 5253.0$ ,  $z = -1.4$ ,  $p = .16$ ,  $r = -0.27$ ) pointed responses. These results suggest that, to the degree tested, the two groups have largely comparable levels of non-verbal emotion understanding, entailing for the hypotheses (specifically, hypothesis groups A and D) that group differences on main tests would not generally be due to differences in 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 100%. In the Naming test, the 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. Hypothesized group differences on main tests are therefore unlikely to be confounded by differences in basic recognition and naming ability of the common items.

### **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 the 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 that the first rater's transcription was reliable. Of the remaining data, Figure 2 summarizes 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 ( $t(1,515) = -3.2$ ,  $p = .001$ ,  $r = 0.69$ ). The NH group was therefore somewhat more accurate at repeating non-words than the CI group. This basic difference has to be taken into account when interpreting results in the main tests, as the NH group might have a somewhat more advanced linguistic age.



**Figure 2.** Mean Percentages of Phonemes Correct per syllable length (in number of syllables) and per participant group (CI or NH) 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.

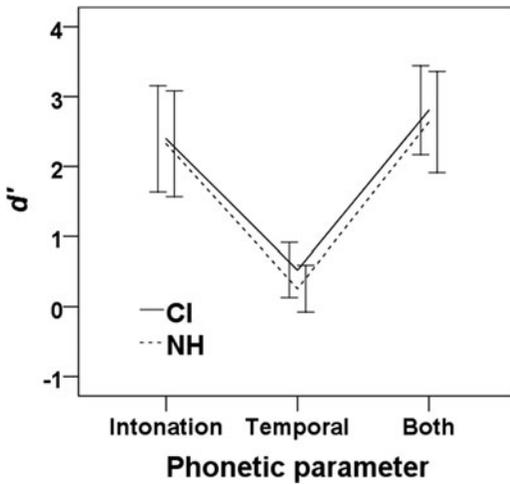
### Main test 1: Emotion perception

Figure 3 shows  $d'$  scores in the Emotion perception test, split by Phonetic parameter (Intonation, Temporal, or Both) and subject group. A mixed-model ANOVA on the  $d'$  scores revealed a main effect of Phonetic parameter ( $F(2,22) = 49.79$ ,  $p < .001$ ), but no effect of Group ( $F(1,23) = 0.18$ ,  $p = .68$ ,  $r = 0.39$ ), nor an interaction between Phonetic parameter and Group ( $F(2,22) = 0.29$ ,  $p = .97$ ). Post-hoc analyses revealed that, of the three Phonetic parameters, scores on the Temporal condition differed highly significantly from both Intonation ( $t(24) = 7.61$ ,  $p < .001$ ,  $r = 0.84$ ) and Both ( $t(25) = -10.70$ ,  $p < .001$ ,  $r = 0.91$ ) conditions, but the Intonation and Both conditions were not significantly different from each other ( $t(24) = -1.79$ ,  $p = .086$ ,  $r = 0.34$ ), whereby a Bonferroni-corrected  $p$ -criterion of  $.05/3$  would have applied if results had approach that critical border. 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. This is in contradiction with hypothesis D.

### Main test 3: Emotion production

Group differences on the Emotion and Focus production tests are not particularly linked to any hypotheses, but the participant-level results are relevant for hypothesis groups A, B, and C, tested as correlations between test scores.

In the Emotion test, of all trials, 3.8% were missing (missing response or technical error). Table 2 and Figure 4 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 ( $F(1,1902) = 7.06$ ,  $p = .008$ ,  $r = 0.061$ ) and 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 fearful ( $F(1,438) = 10.06$ ,  $p = .002$ ,  $r = 0.15$ ) and angry ( $F(1,478) = 14.01$ ,  $p < .001$ ,  $r = 0.17$ ) responses, that the



**Figure 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.

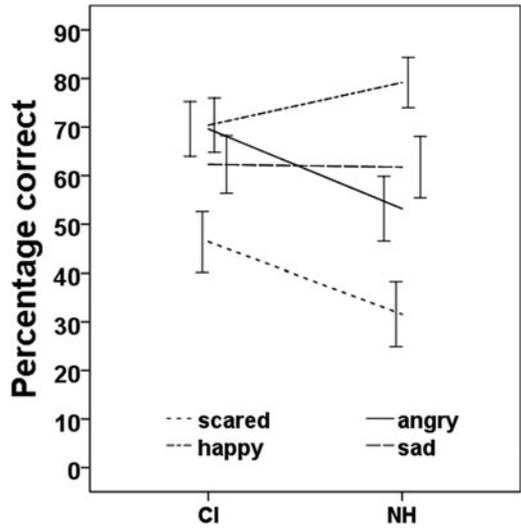
**Table 2.** Mean Percentages Correct and Standard Deviations (in Parentheses) per Emotion and per Participant Group (CI or NH) of Emotions Conveyed in Dummy Phrases in the Emotion Production Test.

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

NH scored better on happy responses ( $F(1,298) = 5.11, p = .024, r = 0.13$ ), but that there was no difference for sad responses ( $F(1,488) = 0.017, p = .90, r = 0.019$ ). These results indicate that the two groups have different specialties when it comes to the production of emotions, but that in general the groups are almost equally good at distinguishing them.

#### Main test 4: Focus production

In the Focus production test, of all trials, 10.9% were missing (missing response or technical error). Table 3 and Figure 5 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 position was found ( $F(2, 2774) = 57.00, p < .001, r = 0.14$ ), but not of Group ( $F(1,2774) = 1.94, p = .164$ ), nor an interaction between Focus position and Group ( $F(2,2774) = 0.94, p = .39$ ). The effect of Focus position was not central to this analysis (which was about group differences), but was conspicuous enough to warrant an explanation. We submit that the score for



**Figure 4.** 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 naive adult Dutch listeners with normal hearing.

**Table 3.** 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	70.0 (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)
<b>Total</b>	<b>58.1 (49.4)</b>	<b>60.4 (48.9)</b>	<b>59.2 (49.2)</b>

adjectives was highest because cases with nuclear accent on the noun were ambiguous between the noun or the whole phrase as a focus domain (Hoekstra, 2000). Therefore, responses for those two options had to be divided between them. The noun option was presumably preferred because of the task-related contrast with the adjective option.

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. These results are relevant for hypothesis groups A, B, and C. 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),

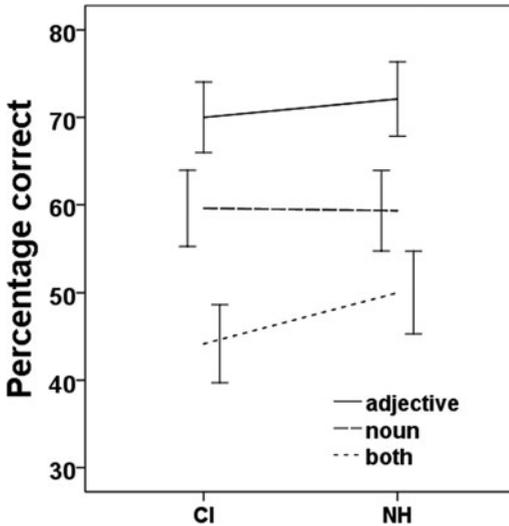


Figure 5. Mean percentages correct per focus position and per participant group (CI or NH) of focus positions conveyed in dummy phrases in the Focus production test. Percentages correct were computed by averaging judgements of emotions perceived by a panel of ten naive adult Dutch listeners with normal hearing.

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 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 normally distributed and that distributions for the other scores were not normally distributed. In the latter case, Kendall's  $\tau$  was reported. It should be noted that for the CI group the distribution of scores on the Non-word repetition test was approaching significance as for the difference from the normal distribution ( $p = .063$ ).

Only the following correlations were significant. For the CI group, scores of the Face discrimination task were marginally significantly and weakly correlated with Emotion perception ( $\tau = 0.387$ ,  $p = .046$ ), those of the Face identification task were moderately correlated with Emotion production ( $\tau = 0.520$ ,  $p = .015$ ), strictly judged verbal responses on the Expression task were moderately or weakly correlated with Emotion perception ( $\tau = 0.453$ ,  $p = .025$ ) and, approaching significance, Focus production ( $\tau = 0.308$ ,  $p = .089$ ), respectively, and strictly judged pointed responses were weakly correlated with Emotion production ( $\tau = 0.398$ ,  $p = .039$ ). In the NH group, strictly judged pointed responses were weakly to moderately and with approaching significance correlated to Focus production ( $\tau = 0.423$ ,  $p = .054$ ). The correlation between Emotion perception and Emotion production was approaching significance for the CI group ( $r = 0.523$ ,  $p = .067$ ), whereas it was not for the NH group ( $r = -0.144$ ,  $p = .656$ ), and the correlation between Non-word repetition and Emotion production was approaching significance for NH group ( $r = 0.543$ ,  $p = .068$ ), whereas it was not significant for the CI group ( $r = 0.017$ ,  $p = .955$ ).

Finally, correlations were run between main test scores on the one hand, and activation age, hearing and chronological age, on the other. In the CI group, the only significant correlation was between hearing age and Emotion production ( $r = 0.542$ ,

$p = .028$ ). In the NH group (where chronological age is by definition equivalent to hearing age), there was a correlation approaching significance between age and Emotion production ( $r = 0.470$ ,  $p = .061$ ).

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 correlating with 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 correlated with 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.

## 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 development, 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 paediatric 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 and others that were not hypothesized. We tested four groups of hypotheses.

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 supported to a limited degree. In the CI group, but not the NH group, Emotion perception performance was moderately and with approaching significance related to 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 go together with better prosody production skills (e.g., Nakata *et al.*, 2012), at least within the domain of emotion. This would suggest that the production of emotions does not 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 *et al.*, 2001). Instead, our results argue in favour of the opposite view, 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 the small sample size warrants caution in interpreting both significant and non-significant effects. 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. The lack of understanding by participants of the focus perception task was not due to a fundamental lack of focus comprehension, as this has been shown to be present to some extent, although with difficulty relative to adults, at this age in previous research (Lee & Snedeker, 2016; Sekerina & Trueswell, 2012). It cannot be ascertained at present if CI children had even more difficulty than NH children, but most likely the task itself was not appropriate. The task problems might have been due to the fact that children needed to respond based on the sound of the stimulus and ignored its contents. This might be more demanding for indexical aspects of speech (as in the emotion perception task) than for linguistic aspects (i.e., focus), whereby the pictures that were the response options bore no meaningful relationship with the target prosody. i.e., a coloured square or line drawing of an object has no inherent link to the concept of emphasis. Moreover, emphasis was realized most clearly on only one word in the stimulus, whereas emotional prosody involved the whole phrase. We submit that this abstraction and this phonetic complication were causes of the lack of results in this task. Future studies should draw attention not to the form but to the meaning of the stimuli by using parallel response options in which either the adjective or the noun contrasts with that of the probe. This would allow for a more straightforward comparison between focus prosody perception and emotion prosody perception competences.

A result not specifically related to a hypothesis was that the CI group produced angry and fearful emotions more effectively than NH controls, but happy phrases less and sad phrases equally successfully, as assessed by raters. This pattern of results might reflect group differences in perceptual acuity of phonetic dimensions. Happiness can be associated with variable pitch (the more difficult perceptual dimension for CI users), whereas anger and fear show variability of loudness (a supposedly easier dimension), as concluded in a literature review (Murray & Arnott, 1993). The correlation approaching significance between perception and production scores for emotion prosody in the current study might have been sufficient for these specializations in emotion production.

As for 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 development, but (B2) that this effect was larger for linguistic than for emotional prosody. As for a relationship with the linguistic baseline test (Non-word repetition), a correlation approaching significance 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, respectively. In the NH group, there was only a correlation that was

approaching significance between Focus production and strictly judged pointed responses in the Expression task. Therefore, Hypothesis B1 (concerning emotion tasks) is supported for the CI group, but not B2, nor either hypothesis for the NH group.

The relative lack of support for hypotheses B1 and B2 and the observation that age effects were somewhat more clearly observed among emotional than among linguistic tests (in the CI group) might imply, in as far as the sample size allows reliance on the results, that perception and production of emotional prosody the way it was tested here is learned to a limited degree. It challenges the theoretically motivated view that emotional prosody competence develops more autonomously than linguistic prosody competence. Apparently, its development, if anything, parallels more general development more closely than linguistic competence does. Although, to our knowledge, this comparison has not yet been directly made in the literature, this result is in line with Wiefferink *et al.* (2013) about the link between emotional development and emotional expression, but not with Filipe *et al.* (2017) about linguistic development. It has to be noted, however, that studies are difficult to compare due to divergent methodologies. Furthermore, NH participants seemed to have already learned to master prosody at the age tested. These learnability results mirror those regarding hypothesis group A, where learnability was speculated to be partial for the experimental but not observed in the control group. However, the results are out of sync with earlier research reporting age effects until around twelve years of age (e.g., Aguert *et al.*, 2013; Filipe *et al.*, 2017). This might be because, whereas the CI children in our study might still be in the process of learning, given that some of their core outcomes did correlate with a subset of developmental measures, for NH children the tasks may not have been cognitively challenging enough to show age differences, in dissonance with some earlier reports (Ito *et al.*, 2014; Wells *et al.*, 2004); indeed, NH children scored better on the Non-word repetition test than the CI children. That is, although the linguistic features tested were probably not fully mastered even by the oldest NH participants (explaining the below-ceiling results), the format of the tasks, usually with two response options and no time limit, was easy for even the youngest NH participants. CI children, on the other hand, might have been delayed in their cognitive development, therefore showing age effects (Burkholder & Pisoni, 2003; however, see Bergeson & Chin, 2008). As more general cognitive abilities were not tested, this account remains open.

This study found no delay in emotional development in CI children. This complements and in part contradicts results by Wiefferink *et al.* (2013), from whose battery of non-verbal emotion understanding tests those in the present study were adopted. They did report such a delay. The lack of a difference in the current study 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 four and eleven years old, it is highly probable that this is due to the fact that the current study tested older children (between six and twelve 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 development 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 development (partly) catches up 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 by Mancini *et al.* (2016) is that for CI children no correlation between performance on verbal emotion

tasks and general developmental 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, was (albeit this was a small effect) sufficient for talking about emotions. That is, they would link them to external referents and learn the synchrony with facial expressions. It has to be noted that CI and NH participants were matched for hearing age (not chronological age) and that 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 development, but (C2) that this effect would be larger for linguistic than for emotional prosody. In the CI group, neither CI activation age nor chronological age was found to be related to outcomes, but hearing age did moderately correlate with Emotion production. In the NH group, age was weakly to moderately, and approaching significance, correlated with Emotion production. The hypotheses are therefore not supported, 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 (the total time before and/or after implantation in which a child experienced auditory input) 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 to improve emotion production performance, again favouring an account of learnability over independent maturation. However, a distinction between perception and production needs to be noted, because the fact that emotion perception was not found to be correlated with this factor would suggest that emotion perception capacities mature independently of hearing experience, whereas emotion production capacities do develop as a function of it. 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 – an account not dictated but left open by correlations – that emotion perception development is a necessary but not a sufficient requirement for emotion production development, whereby one of the other possible requirements is emotional development (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 supported. The recipients weighted their cues in the same way as the children from the control group, namely by relying almost entirely on F0 cues, providing, if reflecting a genuine effect, evidence against 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 that warrant prudence in interpreting the data have to be taken into account. First of all, the samples, with two groups of thirteen participants, were small as a result of the limited availability of implanted children passing the inclusion criteria, compromising generalizability to other paediatric recipients. Some effects and correlations were only approaching significance and although an adopted  $p$  threshold is not to be taken as a strict cut-off between significance and non-significance, such results are to be interpreted with caution.

Second, possibly, the cohort may have 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 for 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 to nuances in emotional development that could influence performance on linguistic prosody tests. More challenging tests, such as those

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 result 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 better perception skills go together with (possibly, allow) better 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 correlated with some measures of non-verbal emotional understanding performance. No such correlation was found for linguistic prosody. For NH children, only contra-modal (from emotion to focus and vice versa) correlations with approaching significance were found. No overall performance level difference was found between groups, suggesting that CI children catch-up with their peers no later than towards the end of primary school.
- (3) Hearing age (itself correlated with activation age for the CI children) weakly correlated with emotion production, but not emotion perception, performance in both groups, suggesting that increasing time in sound has a favourable effect on vocal emotional expression.
- (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 set-up. We also thank all participants for their willingness to cooperate.

### References

- Aguert, M., Laval, V., Lacroix, A., Gil, S., & Le Bigot, L.** (2013). Inferring emotions from speech prosody: not so easy at age five. *PLoS One*, 8(12). doi: 10.1371/journal.pone.0083657
- Baltaxe, C., Simmons, J. Q., & Zee, E.** (1984). Intonation patterns in normal, autistic and aphasic children. In M. P. R. van den Broecke & A. Cohen (Eds.), *Proceedings of the Tenth International Congress of Phonetics Sciences* (pp. 713–18). Dordrecht: Foris.
- Beers, M.** (1995). *The phonology of normally developing and language-impaired children*. Doctoral dissertation, Universiteit van Amsterdam, Amsterdam.
- Bergeson, T., & Chin, S.** (2008). Prosodic utterance production. Manuscript, Indiana University School of Medicine.

- Boons, T., De Raeve, L., Langereis, M., Peeraer, L., Wouters, J., & van Wieringen, A. (2013). Expressive vocabulary, morphology, syntax and narrative skills in profoundly deaf children after early cochlear implantation. *Research in Developmental Disabilities, 34*(6), 2008–22.
- Burkholder, R. A., & Pisoni, D. B. (2003). Speech timing and working memory in profoundly deaf children after cochlear implantation. *Journal of Experimental Child Psychology, 85*(1), 63–88.
- Chen, A. (2007). *Intonational realisation of topic and focus by Dutch-acquiring 4- to 5-year-olds*. Paper presented at the 16th International Congress of Phonetic Sciences, Saarbrücken Retrieved from <[https://pure.mpg.de/rest/items/item\\_57632/component/file\\_57633/content](https://pure.mpg.de/rest/items/item_57632/component/file_57633/content)>.
- Chen, A., Den Os, E., & De Ruiter, J. P. (2007). Pitch accent type matters for online processing of information status: evidence from natural and synthetic speech. *Linguistic Review, 24*(2/3), 317–44.
- Chin, S. B., Tsai, P. L., & Gao, S. (2003). Connected speech intelligibility of children with cochlear implants and children with normal hearing. *American Journal of Speech-Language Pathology, 12*(4), 440–51.
- Clark, A. (2007). *Acoustic correlates of linguistic prosody in the speech of children with cochlear implants: a study in comparison with typical-hearing peers*. Doctoral dissertation, University of Colorado, Boulder.
- Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology, 65*(2), 171–237.
- De Bree, E., Rispens, J., & Gerrits, E. (2007). Non-word repetition in Dutch children with (a risk of) dyslexia and SLI. *Clinical Linguistics & Phonetics, 21*(11/12), 935–44.
- de Ruiter, L. (2010). *Studies on intonation and information structure in child and adult German*. Doctoral dissertation, Radboud Universiteit Nijmegen.
- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41*(5), 1136–46.
- Dorman, M. F., & Loizou, P. C. (1998). The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels. *Ear and Hearing, 19*(2), 162–66.
- Filipe, M. G., Peppé, S., Frota, S., & Vicente, S. G. (2017). Prosodic development in European Portuguese from childhood to adulthood. *Applied Psycholinguistics, 38*(5), 1045–70.
- Flom, R., & Bahrack, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: the role of intersensory redundancy. *Developmental Psychology, 43*(1), 238–52.
- Fu, Q. J., Chinchilla, S., Nogaki, G., & Galvin, J. J. (2005). Voice gender identification by cochlear implant users: the role of spectral and temporal resolution. *Journal of the Acoustical Society of America, 118*(3), 1711–18.
- Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q. J., Free, R. H., & Baskent, D. (2014). Gender categorization is abnormal in cochlear implant users. *Journal of the Association for Research in Otolaryngology, 15*(6), 1037–48.
- Garrapa, L. (2014). *Vowel processing in Italian pediatric cochlear-implant users: a behavioral and neurophysiological study*. Doctoral dissertation, University of Padua.
- Geers, A., Davidson, L. S., Uchanski, R. M., & Nicholas, J. G. (2013). Interdependence of linguistic and indexical speech perception skills in school-age children with early cochlear implantation. *Ear and Hearing, 34*(5), 562–74.
- Geers, A., Nicholas, J., Tobey, E., & Davidson, L. (2016). Persistent language delay versus late language emergence in children with early cochlear implantation. *Journal of Speech, Language, and Hearing Research, 59*(1), 155–70.
- Geers, A., Nicholas, J. G., & Sedey, A. L. (2003). Language skills of children with early cochlear implantation. *Ear and Hearing, 24*(1), 46s–58s.
- Geers, A., Tobey, E., Moog, J., & Brenner, C. (2008). Long-term outcomes of cochlear implantation in the preschool years: from elementary grades to high school. *International Journal of Audiology, 47*, S21–S30.
- Gfeller, K., Turner, C., Oleson, J., Zhang, X. Y., Gantz, B., Froman, R., & Olszewski, C. (2007). Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. *Ear and Hearing, 28*(3), 412–23.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*: Cambridge University Press.
- Han, D., Zhou, N., Li, Y., Chen, X., Zhao, X., & Xu, L. (2007). Tone production of Mandarin Chinese speaking children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology, 71*(6), 875–80.

- Helms, J., Müller, J., Schön, F., Moser, L., Arnold, W., Janssen, T., ... Pfennigdorf, T.** (1997). Evaluation of performance with the COMBI 40 cochlear implant in adults: a multicentric clinical study. *ORL*, 59(1), 23–35.
- Hockett, C. F.** (1960). The origin of speech. *Scientific American*, 203(3), 88–97.
- Hoekstra, H.** (2000). An algorithm for the assignment of sentence accents in Dutch. *Linguistics in the Netherlands*, 17(1), 105–18.
- Holt, C. M.** (2013). *The perception and production of prosody by adolescent users of cochlear implants*. Doctoral dissertation, University of Melbourne.
- Hopyan-Misakyan, T. M., Gordon, K. A., Dennis, M., & Papsin, B. C.** (2009). Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. *Child Neuropsychology*, 15(2), 136–46.
- Horst, J. S., & Hout, M. C.** (2016). The Novel Object and Unusual Name (NOUN) database: a collection of novel images for use in experimental research. *Behavior Research Methods*, 48(4), 1393–1409.
- IBM** (2012). *IBM SPSS Statistics for Windows, Version 23.0* [computer program]. Armonk, NY: IBM Corp.
- Ito, K., Bibyk, S. A., Wagner, L., & Speer, S. R.** (2014). Interpretation of contrastive pitch accent in six- to eleven-year-old English-speaking children (and adults). *Journal of Child Language*, 41(1), 84–110.
- Iverson, P., Smith, C. A., & Evans, B. G.** (2006). Vowel recognition via cochlear implants and noise vocoders: effects of formant movement and duration. *Journal of the Acoustical Society of America*, 120(6), 3998–4006.
- Johnson, E. K., & Jusczyk, P. W.** (2001). Word segmentation by 8-month-olds: when speech cues count more than statistics. *Journal of Memory and Language*, 44(4), 548–67.
- Kord, N., Shahbodaghi, M. R., Khodami, S. M., Nourbakhsh, M., & Jalaei, S.** (2013). Investigation of acoustic correlation of intonation and intelligibility of speech in children with cochlear implant and comparison with normal hearing children. *Modern Rehabilitation*, 6(4), 38–43.
- Kraayeveld, H.** (1997). *Idiosyncrasy in prosody: speaker and speaker group identification in Dutch using melodic and temporal information*. Doctoral dissertation, University of Nijmegen.
- Lee, E.-K., & Snedeker, J.** (2016). Effects of contrastive accents on children's discourse comprehension. *Psychonomic Bulletin & Review*, 23(5), 1589–95.
- Lenden, J. M., & Flipsen, P.** (2007). Prosody and voice characteristics of children with cochlear implants. *Journal of Communication Disorders*, 40(1), 66–81.
- Liu, H. M., Kuhl, P. K., & Tsao, F. M.** (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, 6(3), F1–F10.
- Luo, X., Fu, Q. J., & Galvin, J. J.** (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification*, 11(4), 301–15.
- Lyxell, B., Wass, M., Sahlen, B., Samuelsson, C., Asker-Arnason, L., Ibertsson, T., ... Hallgren, M.** (2009). Cognitive development, reading and prosodic skills in children with cochlear implants. *Scandinavian Journal of Psychology*, 50(5), 463–74.
- Macmillan, N. A., & Kaplan, H. L.** (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, 98(1), 185–99.
- Macmillan, N. A., & Creelman, C. D.** (2004). *Detection theory: a user's guide*. Mahwah, NJ/London: Lawrence Erlbaum Associates.
- Mancini, P., Giallini, I., Prosperini, L., D'Alessandro, H. D., Guerzoni, L., Murri, A., ... Nicastrì, M.** (2016). Level of emotion comprehension in children with mid to long term cochlear implant use: how basic and more complex emotion recognition relates to language and age at implantation. *International Journal of Pediatric Otorhinolaryngology*, 87, 219–32.
- Meister, H., Landwehr, M., Pyschny, V., Wagner, P., & Walger, M.** (2011). The perception of sentence stress in cochlear implant recipients. *Ear and Hearing*, 32(4), 459–67.
- Meister, H., Landwehr, M., Pyschny, V., Walger, M., & von Wedel, H.** (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *International Journal of Audiology*, 48(1), 38–48.
- Meister, H., Tepeli, D., Wagner, P., Hess, W., Walger, M., von Wedel, H., & Lang-Roth, R.** (2007). [Experiments on prosody perception with cochlear implants]. *HNO*, 55(4), 264–70. doi: 10.1007/s00106-006-1452-1
- Most, T., & Peled, M.** (2007). Perception of suprasegmental features of speech by children with cochlear implants and children with hearing AIDS. *Journal of Deaf Studies and Deaf Education*, 12(3), 350–61.

- Murray, I. R., & Arnott, J. L.** (1993). Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, 93(2), 1097–108.
- Nakata, T., Trehub, S. E., & Kanda, Y.** (2012). Effect of cochlear implants on children's perception and production of speech prosody. *Journal of the Acoustical Society of America*, 131(2), 1307–14.
- Neuman, A. C.** (2014). Reverberation and noise pose challenges to speech recognition by cochlear implant users. *Journal of the Acoustical Society of America*, 135(4), 2365.
- O'Halpin, R.** (2009). *The perception and production of stress and intonation by children with cochlear implants*. Doctoral dissertation, UCL, London. Retrieved from <<http://eprints.ucl.ac.uk/20406/1/20406.pdf>>.
- Osberger, M. J.** (1994). Speech intelligibility of children with cochlear implants. *Volta Review*, 96(5), 169–80.
- Peng, S. C., Lu, N., & Chatterjee, M.** (2009). Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners. *Audiology and Neuro-Otology*, 14(5), 327–37.
- 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 and Hearing*, 29(3), 336–51.
- Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M.** (2007). Receptive and expressive prosodic ability in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research*, 50(4), 1015–28.
- Pierrehumbert, J.** (1980). *The phonetics and phonology of English intonation*. Doctoral dissertation, MIT, Cambridge, MA.
- Rietveld, A. C. M., & Van Heuven, V. J.** (2016). *Algemene fonetiek*, 4th ed. Bussum: Coutinho.
- Robinson, K.** (1998). Implications of developmental plasticity for the language acquisition of deaf children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, 46(1), 71–80.
- Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P.** (2006). Vocal expression of emotions in normally hearing and hearing-impaired infants. *Journal of Voice*, 20(4), 585–604.
- Scherer, K. R., Banse, R., & Wallbott, H. G.** (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-cultural Psychology*, 32(1), 76–92.
- Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T.** (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, 15(2), 123–48.
- Schiller, N. O.** (2006). Lexical stress encoding in single word production estimated by event-related brain potentials. *Brain Research*, 1112(1), 201–12.
- Schlichting, J. E. P. T., & Lutje Spelberg, H. C.** (2002). *Lexilijst Nederlands: een instrument om de taalontwikkeling te onderzoeken bij Nederlandstalige kinderen van 15-27 maanden in het kader van de vroegonderkenning: Manual*. Lisse: Swets Test Publishers.
- Schneider, W., Eschman, A., & Zuccolotto, A.** (2012). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools, Inc.
- Sekerina, I. A., & Trueswell, J. C.** (2012). Interactive processing of contrastive expressions by Russian children. *First language*, 32(1/2), 63–87.
- Shriberg, L. D., & Kwiatkowski, J.** (1994). Developmental phonological disorders. I: a clinical profile. *Journal of Speech and Hearing Research*, 37(5), 1100–26.
- Sityaev, D., & House, J.** (2003). *Phonetic and phonological correlates of broad, narrow and contrastive focus in English*. Paper presented at the Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain Retrieved from <[https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15\\_1819.pdf](https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15_1819.pdf)>.
- Souza, L. B., Bevilacqua, M. C., Brasolotto, A. G., & Coelho, A. C.** (2012). Cochlear implanted children present vocal parameters within normal standards. *International Journal of Otorhinolaryngology*, 76(8), 1180–3.
- Stanislaw, H., & Todorov, N.** (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137–49.
- Straatman, L. V., Rietveld, A. C. M., Beijen, J., Mylanus, E. A. M., & Mens, L. H. M.** (2010). Advantage of bimodal fitting in prosody perception for children using a cochlear implant and a hearing aid. *Journal of the Acoustical Society of America*, 128(4), 1884–95.
- Ubrig, M. T., Goffi-Gomez, M. V., Weber, R., Menezes, M. H., Nemr, N. K., Tsuji, D. H., & Tsuji, R. K.** (2011). Voice analysis of postlingually deaf adults pre- and postcochlear implantation. *Journal of Voice*, 25(6), 692–9.
- Vällimaa, T. T., Määttä, T. K., Löppönen, H. J., & Sorri, M. J.** (2002). Phoneme recognition and confusions with multichannel cochlear implants: vowels. *Journal of Speech, Language, and Hearing Research*, 45(5), 1039–54.

- Välilmaa, T. T., Sorri, M. J., Laitakari, J., Sivonen, V., & Muhli, A. (2011). Vowel confusion patterns in adults during initial 4 years of implant use. *Clinical Linguistics & Phonetics*, 25(2), 121–44.
- van de Velde, D. J., Schiller, N. O., van Heuven, V. J., Levelt, C. C., van Ginkel, J., Beers, M., ... Frijns, J. H. (2017). The perception of emotion and focus prosody with varying acoustic cues in cochlear implant simulations with varying filter slopes. *Journal of the Acoustical Society of America*, 141(5), 3349–63.
- van Heuven, V. J. (1992). Linguistic versus phonetic explanation of consonant lengthening after short vowels; a contrastive study of Dutch and English. In *Proceedings of the 2nd International Conference on Spoken Language Processing* (pp. 1275–7), Banff Retrieved from <[https://www.isca-speech.org/archive/archive\\_papers/icslp\\_1992/i92\\_1275.pdf](https://www.isca-speech.org/archive/archive_papers/icslp_1992/i92_1275.pdf)>.
- Van Lierde, K. M., Vinck, B. M., Baudonck, N., De Vel, E., & Dhooge, I. (2005). Comparison of the overall intelligibility, articulation, resonance, and voice characteristics between children using cochlear implants and those using bilateral hearing aids: a pilot study. *International Journal of Audiology*, 44(8), 452–65.
- Volkova, A., Trehub, S. E., Schellenberg, E. G., Papsin, B. C., & Gordon, K. A. (2013). Children with bilateral cochlear implants identify emotion in speech and music. *Cochlear Implants International*, 14(2), 80–91.
- Wang, D., Trehub, S., Volkova, A., & van Lieshout, P. (2013). Child implant users' imitation of happy- and sad-sounding speech. *Frontiers in Psychology*, 4, 351. doi.org/10.3389/fpsyg.2013.00351
- Wells, B., Peppe, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language*, 31(4), 749–78.
- Wiefferink, C. H., de Vries, M., & Ketelaar, L. (2015). From fundamental research to a functional tool for professionals. In H. Knoors & M. Marschark (Eds.), *Educating deaf learners: creating a global evidence base* (pp. 461–80). New York: Oxford University Press.
- Wiefferink, C. H., Rieffe, C., Ketelaar, L., De Raeve, L., & Frijns, J. H. M. (2013). Emotion understanding in deaf children with a cochlear implant. *Journal of Deaf Studies and Deaf Education*, 18(2), 175–86.
- Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: some acoustical correlates. *Journal of the Acoustical Society of America*, 52(4), 1238. doi: 10.1121/1.1913238
- Witteman, J., van IJzendoorn, M. H., van de Velde, D., van Heuven, V. J. J. P., & Schiller, N. O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: a meta-analysis of the lesion literature. *Neuropsychologia*, 49(13), 3722–38.

## Appendix

The stimuli used in the non-word repetition task. No experimental items functioned as practice items. There were no practice items consisting of one, two, or four syllables.

1 syllable	2 syllables	3 syllables	4 syllables	5 syllables
Practice	Practice	Practice	Practice	Practice
–	–	Nietoofoes	–	Piefoesaanooteem
<b>Experimental</b>	<b>Experimental</b>	<b>Experimental</b>	<b>Experimental</b>	<b>Experimental</b>
Noos	Noetiem	Sootienoem	Naafiesooteem	Taanoosoefiepeem
Fiem	Saapoom	Taapienoes	Siepootafoem	Saanoepootiefem
Taas	Tienees	Piefoenoom	Foepietoonees	Faanietoosoepeem
Poem	Pietoes	Faapooties	Poofienaatees	Niepoofaanoetees

**Cite this article:** Van De Velde DJ, Schiller NO, Levelt CC, Van Heuven VJ, Beers M, Briaire JJ, Frijns JHM (2019). Prosody perception and production by children with cochlear implants. *Journal of Child Language* 46, 111–141. <https://doi.org/10.1017/S0305000918000387>