Expansion of Prosodic Abilities at the Transition from Babble to Words: a Comparison Between Children with Cochlear Implants and Normally Hearing Children[1]

Michèle Pettinato[a], Ilke De Clerck[a], Jo Verhoeven[a,b], Steven Gillis[a]

[a] University of Antwerp, Department of Linguistics, CLiPS Computational Linguistics and Psycholinguistics research centre, Antwerp, Belgium.

[b] City University London, Division of Language and Communication Science, London, UK.

Corresponding author:

Michèle Pettinato

University of Antwerp

Department of Linguistics, CLiPS

Prinsstraat 13

2000 Antwerp

Belgium

Tel. +3232655239

Email michele.pettinato@uantwerpen.be

[1] An earlier version of this work was presented in "The production of word stress in babbles and early words: a comparison between normally hearing infants and infants with cochlear implants" at the 18th International Congress of Phonetic Sciences (ICPhS), Glasgow,10 –14th August 2015

Conflicts of Interest and Source of Funding:

This research was funded by a PhD Fellowship grant of the Research Foundation – Flanders (FWO) and a BOF-DOCPRO project (ID 28259) from the Research Council of the University of Antwerp.

The authors declare that there are no conflicts of interest.

Acknowledgements:

We thank the editors and two reviewers for their comments which substantially improved this manuscript.

ABSTRACT

Objectives: This longitudinal study examined the impact of emerging vocabulary production on the ability to produce the phonetic cues to prosodic prominence in babbled and lexical disyllables of infants with Cochlear Implants (CI) and normally hearing infants (NH). Current research on typical language acquisition emphasizes the importance of vocabulary development for phonological and phonetic acquisition. Children with cochlear implants (CI) experience significant difficulties with the perception and production of prosody, and the role of possible top-down effects is therefore particularly relevant for this population.

Design: Isolated disyllabic babble and first words were identified and segmented in longitudinal audio-video recordings and transcriptions for 9 NH infants and 9 infants with CI interacting with their parents. Monthly recordings were included from the onset of babbling until children had reached a cumulative vocabulary of 200 words. Three cues to prosodic prominence, F0, intensity and duration, were measured in the vocalic portions of stand-alone disyllables. In order to represent the degree of prosodic differentiation between two syllables in an utterance, the raw values for intensity and duration were transformed to ratios, and for f0 a measure of the perceptual distance in semitones was derived. The degree of prosodic differentiation for disyllabic babble and words for each cue was compared between groups. In addition, group and individual tendencies on the types of stress patterns for babble and words were also examined.

Results: The CI group had overall smaller pitch and intensity distances than the NH group. For the NH group, words had greater pitch and intensity distances than babbled disyllables. Especially for pitch distance, this was accompanied by a shift towards a more clearly expressed stress pattern that reflected the influence of the ambient language. For the CI group, the same expansion in words did not take place for pitch. For intensity, the CI group gave evidence of some increase of prosodic differentiation. The results for the duration measure showed evidence of utterance-final lengthening in both groups. In words, the CI group significantly reduced durational differences between syllables so that a more even-timed, less differentiated pattern emerged.

Conclusions: The onset of vocabulary production did not have the same facilitatory effect for the CI infants on the production of phonetic cues for prosody, especially for pitch. It was argued that the results for duration may reflect greater articulatory difficulties in words for the CI group than the NH group. It was suggested that the lack of clear top-down effects of the vocabulary in the CI group may be due to a lag in development caused by an initial lack of auditory stimulation, possibly compounded by the absence of auditory feedback during the babble phase.

INTRODUCTION

The role of the lexicon in phonological and phonetic acquisition has increasingly been emphasised over the last decade or so (Pierrehumbert, 2003; Stoel-Gammon, 2011). Whereas previous research tended to take a bottom-up approach which concentrated on distributional properties of the acoustic input (Maye, Werker, & Gerken, 2002) or structural-linguistic influences on development (Dinnsen, Green, Morrisette, & Gierut, 2011), there has been a resurgence of interest in the highlighting and constraining effect of vocabulary development on speech perception and production. The present study considers the effect of the emergent vocabulary production on the acoustic phonetics of prosody in early disyllabic utterances of normally hearing (henceforth NH) and severely to profoundly hearing impaired children with cochlear implants (henceforth CI). Prosody was deemed a particular area of interest because it is highly important in typical language acquisition (Morgan & Demuth, 1996), but the implant only provides a significantly degraded input for two of the acoustic cues to prosody (Moore, 2003). For children with CI, an exploration of top-down processes on language acquisition is particularly relevant, as these could in principle be used to counter the inherent limitations in signal processing of the implant. Indeed, recent theories of language development have emphasised bi-directional relations between earlier more basic processes and the acquisition of later, higher-order abilities, where the first not only constrains the second but is in turn shaped by it (Werker & Tees, 2005). Exploring the effect of vocabulary acquisition on phonetic development in children with CI is therefore timely, as we do not know whether children with CI are able to draw the same benefit from vocabulary learning as NH children.

The Role Of The Lexicon In Phonological and Phonetic Acquisition

In the earliest stages of language acquisition, the lexicon may already guide infants' phonetic perception. Computational simulations showed that a small proto-lexicon can make boundaries between phonetic categories clearer (Feldman, Griffiths, and Morgan, 2009). Indeed, having a speech sound presented in a word can make it easier for infants to learn to distinguish it (Yeung & Werker, 2009). For speech production, children's phonological and phonetic abilities have been linked to the size of their vocabularies (Beckman & Edwards, 2000; Nicholson, Munson, Reidy, & Edwards, 2015), and even the stability of articulatory speech movements is greater when nonsense words designate a referent, rather than being repeated as a mere string of syllables (Heisler, Goffman, & Younger, 2010). Vocabulary effects have not only been described for segmental acquisition, but also for prosody. Work by DePaolis, Vihman and Kunnari (2008) suggests reliable word stress may not emerge until first words begin to be produced. 'Word stress' refers to differences in audible syllable prominence in a word. In acoustic terms, stressed syllables, e.g. the syllable 'ni' in 'vanilla', typically have higher f0, intensity and longer duration (Lieberman, 1960; Kochanski, Grabe, Coleman, & Rosner, 2005).

Recently, we have argued for an enhancing role of vocabulary in the acquisition of word stress for typically developing children (De Clerck, Pettinato, Verhoeven & Gillis, in press). We showed a dramatic expansion of prosodic differentiation in first words. Monthly recordings from nine typically developing Belgian-Dutch speaking infants were analysed from the onset of babbling until a cumulative vocabulary of 200 words was reached. The majority of disyllabic lexical forms in Dutch start with a stressed syllable, giving rise to a trochaic (strong-weak) pattern (Cutler, 2005; Daelemans, Gillis, & Durieux, 1994). This pattern was already visible in the babbled utterances of infants but became much clearer in first words: words showed significantly more prosodic differentiation in terms of f0 and intensity. This effect was robust to individual variation. The increase took place abruptly as soon as first words appeared and did not seem to relate to the gradual increase in vocabulary size. It was argued that this was because the advent of words brought about increased attention and allocation of resources to phonetic detail.

These results may be important because prosody, and in particular word stress, has been assigned a critical role in language acquisition over the past three decades (Morgan & Demuth, 1996). Word stress is perceived early in development (Friederici, Friedrich & Christophe, 2007), and children can use this information to detect words in the continuous speech stream and as an entry point to the syntax of their language (Jusczyk, Houston, & Newsome, 1999; Curtin, Campbell, & Hufnagle, 2012). Moreover, children's early word forms and syllable omissions show a strong influence of the most frequent stress patterns in the ambient language (Demuth, 1996), and prosodic constraints on the development of morpho-syntactic production abilities have also been described (Gerken, 1994).

Difficulties With The Perception And Production Of Prosody In Children With CI

Given the suggested importance of word stress, a pertinent question is what happens in language acquisition when infants do not have easy access to the phonetic cues of word stress. This is the case for children with cochlear implants. The spectral and temporal resolution of the implant does not afford enough detail for adequate f0 perception (Moore, 2003; Green, Faulkner, & Rosen, 2004; O'Halpin, 2010) or changes in intensity (Drennan & Rubinstein, 2008; Moore, 2003; Meister, Landwehr, Pyschny, Wagner, and Walger, 2011), but durational properties of syllables seem to be available to listeners with a CI (O'Halpin, 2010; Meister et al., 2011). Whilst f0 is not available as a cue for prosody, temporal aspects of the amplitude envelope may still provide cues to pitch (Green et al., 2004; O'Halpin, 2010).

In children and adults with CI, impaired perception has indeed been reported for word and sentence stress (Most & Peled, 2007; O'Halpin, 2010; Titterington, Henry, Krämer, Toner, & Stevenson, 2006). Adult and child listeners with CI experience significant difficulties when asked to determine the emotion behind an utterance or whether it is a question or a statement (Hopyan-Misakyan, Gordon, Dennis, & Papsin, 2009; Nakata, Trehub, & Kanda, 2012; Peng, Tomblin, & Turner, 2008). Indeed, if listeners with CI succeed in f0 shape and alignment perception, they do so on far less fine-grained distinctions (Holt & McDermott, 2013; Holt & Fletcher 2015). However, some of the negative findings for prosody perception may have been influenced by the relatively late age of implantation of the participants in most of the studies reviewed. This is suggested by Torppa et al. (2014), who report equivalent perception of prosody (word and sentence stress) to NH age-matched peers in early-implanted (before three years), musically trained school-aged children with CI. Furthermore, recent evidence of Hebrew-acquiring infants with CI (chronological ages 13-33 months) suggests that early implanted infants may develop a similar, although less pronounced, sensitivity to the predominant stress pattern of their native language as NH infants (Segal, Houston, & Kishon-Rabin, 2015).

For production, Lenden and Flipsen (2007) noted abnormalities in word and sentence stress in a study of conversational speech samples of six children with CI (chronological ages 3-6 years). Stress production sounded 'excessive, equal or misplaced' (Lenden & Flipsen, 2007, p.75), whilst measures of phrasing, voice quality and pitch were relatively unaffected. On nonsense word repetition tasks, 8-9 year old children with CI only reached 61% accuracy for stress placement (Carter, Dillon & Pisoni, 2002). In contrast, a study with 6-9 year old Belgian Dutch-speaking children with CI, most of whom had been implanted before the age of 2 years, found that children's nonsense word repetitions were mostly rated correctly stressed by adult listeners (Hide, 2013). Nevertheless, acoustic measurements revealed that the children with CI made less distinct acoustic differences than a NH control group of the same age. To summarise, the acquisition of word stress is likely to be more effortful, though not impossible, for children with CI as it relies on acoustic distinctions which are harder to perceive and more fragile to noise perturbation (Peters, Moore, & Baer, 1998). The acoustic cues are unlikely to be accessible to all CI listeners, though the ability to learn to use the reduced information transmitted by the implant may be partly dependent on early implantation.

An Exploration Of The Role Of Vocabulary Development on the production of phonetic cues to prosody in children with CI

The main aim of the present study was therefore to find out if the same increase in the ability to produce the phonetic cues to prosody occurs on the first words of children with CI as in NH children. In order to investigate this question, prosodic modulation in disyllabic babble and first words of 9 children with CI was compared to that of the 9 NH children described in De Clerck et al. (in press). Fundamental frequency (f0), intensity and duration were measured in the vocalic portions of disyllabic utterances of infants acquiring Belgian Dutch. A secondary aim was to compare developmental trajectories of the acoustic cues in the two groups.

Hide's (2013) results with early-implanted children with CI led us to hypothesise that the CI group should be able to converge towards the same prosodic pattern as the NH group, but it was not clear how early this would emerge, nor how robust this would be. Because of the insufficient signal processing of the implant, it was also expected that the CI group should show reduced use of f0 and potentially intensity, but should not differ from the NH group in the use of duration. The prediction regarding the impact of first words was less clear: on the one hand, top-down effects from the vocabulary had served to enhance prosodic-phonetic development in NH children (de Clerck et al., in press), and should therefore also be expected to strengthen development in the CI group. On the other hand, the initial absence of stimulation raises the possibility of sensitive periods being disrupted (Knudsen, 2004; Sharma, Dorman & Kral, 2005) and the CI group has overall had less exposure to speech, meaning that the arrival of first words may not be enough to trigger greater prosodic differentiation of phonetic cues in this group. To summarise, the following research questions were investigated in this study:

1. Is the onset of vocabulary development accompanied by the same expansion of prosodic modulation in infants with CI as in NH infants?
2. Do the two groups follow comparable developmental trajectories? I.e. does the development of cue use reflect the limitations of the implant, and do the groups converge on similar prosodic patterns?

MATERIALS AND METHODS

Participants

The data for this study were taken from the CLiPS Child Language Corpus (CCLC), a collection of longitudinal audio-video data and transcriptions of 10 children with a cochlear implant (CI) and 40 normally hearing children (NH). All parents of the children in the CCLC had signed an informed consent form.

For the purposes of the present study 9 children with a cochlear implant were included: one participant had to be excluded because there were not enough recordings to yield a sufficient number of data points. The children with CI were recruited from an Academic ENT Unit in Antwerp/Belgium in 2000-2001. These participants had all been diagnosed with a profound congenital hearing loss on the basis of a neonatal hearing screening during the first weeks of life. No other patent health or developmental problems were reported. All these children had been implanted with a multichannel Nucleus-24 CI (Cochlear Corp., Sydney, Australia). The Nucleus-24 device consists of 22 intra-cochlear electrodes, like more recent CIs. Since the technology of the implant has not changed in such a way that the perception of fundamental frequency has significantly improved, the data in the present study are still representative.

The infants were implanted before the age of two, ranging from 6 to 19 months (*M* = 12 months; *SD* = 5 months). The average unaided hearing loss was more than 90 dBHL in the better ear. Before implantation the range of the Pure Tone Averages (PTA) was 93-120 dBHL (*M* = 113 dBHL; *SD* = 9 dB). After implantation, as measured around one year after fitting of the implant, the PTA decreased to 30-52 dBHL (*M* = 40 dBHL; *SD* = 7 dB). All recordings used in this study were made while the children were unilaterally implanted. Only CI-7 received her second CI in the same month as the last recording used. The auditory characteristics of the children with CI are summarized in Table 1. For more information on the participants and the aetiology of their hearing losses, see Schauwers, Gillis and Govaerts (2008).

INSERT TABLE 1 ABOUT HERE

The mean age of the children with CI at the start of the recordings was 17 months (*SD* = 4 months). The mean age at the cut-off point was 24 months (*SD* = 4 months). The ages of the individual children at the time of recording are summarized in Table 2.

INSERT TABLE 2 ABOUT HERE

As a control group for this study, 9 normally-hearing children were selected from the CCLC corpus. All families from both the CI and NH groups are considered to be from mid to high socio-economic class, based on the parents' education, wage and occupation. At least one parent had a bachelor or master degree (80% of all parents had a bachelor or masters degree), the income level was above the minimum wage and all parents worked full time. The infants were recruited from day-care centres, families known by the researchers and by announcements. Just like the children with CI, the normally-hearing children had been raised in monolingual homes acquiring the standard variety of Belgian Dutch (Verhoeven, 2005). The typical development of these children had been established on the basis of parent report and a checklist of the attainment of communicative and motor milestones, largely based on the checklist developed by 'Kind en Gezin', the Belgian infant welfare centre (Molemans, van den Berg, van Severen, & Gillis, 2012). Normal language development had been verified by means of the Dutch version of the CDI ("N-CDI") administered at ages (years; months) 1;0, 1;6 and 2;0 (Zink & Lejaegere, 2001). The N-CDI was filled out by the parents of the NH children to test productive and receptive vocabulary development. The mean percentile for the infants included in this study was 37,9 (SD = 28,4; range = 5.5 - 94.5) at 1;0, 46.9 (SD = 23; range = 20-90) at 1;6 and 51.7 (SD = 29.5; range = 10-90) at 2;0. The mean age of the NH children at the time of the recordings was 6 months (SD = 0,72 months). The mean age at the cut-off point was 22 months (SD = 3 months). The ages of the individual children at the time of recording are specified in Table 2.

Corpus

The corpus consisted of monthly recordings of spontaneous interactions between the children and their caretakers in their home environment (for more details on the corpus and transcription, see: Molemans et al., 2012). A JVC digital video was used to record the NH participants while children with a CI were filmed with a Panasonic NV-GS3 digital video camera with zoom microphone function. Recordings lasted 60-90 minutes and the fragments during which the child was most vocally active and in uninterrupted interaction with a caretaker were selected. The final selection was 20 minutes long.

These interactions were transcribed following CHILDES CHAT conventions (MacWhinney, 2000). The criteria for distinguishing words from babble were based on Vihman and McCune (1994). In order to qualify as a lexical item, utterances had to meet at least two out of three criteria: a determining context or the mother's identification clarified the meaning (e.g. the child utterance *baba* was interpreted as *bal* ‘ball’ by the mother), an exact or prosodic match to the target word (i.e. *pal* for *bal* ‘ball’), or the relation to other vocalisations such as imitation or an invariant production (i.e. consistent use of *popo* for *opa* ‘grandpa’).

CHAT transcriptions were converted to a Praat (Boersma & Weenink, 2014) texgrid using the CHAT2PRAAT function in the CLAN program (MacWhinney, 2013). The video files were converted to audio files by means of Free-Video-Converter ("Free-Video-Converter," 2012). The resulting textgrids were time-aligned at the utterance level to the audio files as illustrated in Figure 1.

Data selection

For the present study, speech data were included from the onset of babbling until children had reached a cumulative vocabulary of 200 words. This cut-off point was randomly chosen but motivated by other studies using vocabulary level as developmental point (Vihman, DePaolis & Davis, 1998). Onset of babbling was determined by a True Canonical Babbling Ratio (tCBR) of 0.15 or higher (Chapman, Hardin-Jones, Schulte, & Halter, 2001; Molemans et al., 2012). The tCBR is the proportion of the syllables with true consonants (i.e. all consonants except glottals (/h/, glotal stop) and glides (/w/, /j/)) over all syllables produced. Cumulative vocabulary was used as a measurement of lexical development and was obtained by counting the different word types produced per transcribed recording. The newly produced word types in the following recording were added to the amount of different word types of previous recordings and so forth. The cut-off point of a cumulative vocabulary of 200 words was motivated by the amount of data that was provided in the recordings. The aim was to incorporate enough data to sketch a substantial profile of the longitudinal development of prominence production. Since a longitudinal approach is taken in the current study, no artificial boundaries are placed between a babbling phase and a lexical phase as there is a transitional period where babble and words co-occur. Disyllables tagged as babble are likely to contain a number of words, and words are likely to contain some babble, especially during the transition phase when these co-occur. To illustrate: the status of a particular utterance may be unclear as it may start out as babble, but be interpreted as a word by the parents. Conversely, attempts at words may not be recognized as such. Both scenarios are equally likely, thus this noise should be evened out statistically. The fact that this is a longitudinal study, which goes beyond this transitional phase also serves to counteract this temporary noise.

Inclusion Criteria For Disyllables.

The waveforms and spectrograms associated with the speech files were examined in order to identify the words and babble so that they could be tagged in the PRAAT textgrids (see Figure 1). Sound sequences were considered to be disyllables when they consisted of two vocalic phases minimally separated by a clear consonantal phase (see *Segmentation criteria for consonants and vowels* for specification). Additional consonants flanking the vocalic sections were allowed. The inclusion criteria for the disyllables were based on DePaolis et al. (2008). In order to be included, disyllables had to be clearly perceived as single utterance. This meant that the two syllables of the utterance had to occur within the same intonation contour or adjacent to a prosodic break such as a pause or an inbreath at the beginning of a new breath group (Lieberman, 1984). Furthermore, disyllables had to be separated from surrounding speech by a silence of at least 400 ms with an intersyllabic pause of less than 400 ms. For a small number of items produced at a low speech rate, the pause criterion was relaxed up to 500ms as long as the two syllables were part of a single intonation contour, indicating cohesion. Disyllables were excluded if there was concurrent speech or noise or if they were produced with a creaky, breathy, excessively loud or whispery voice. An example of a selected utterance is provided in Figure 1.

Segmentation Criteria For Consonants And Vowels.

The disyllables identified by the procedure described above were further segmented into consonants and vowels, since the acoustic measurements in this study were conducted in the vocalic portion of each syllable. Figure 1 illustrates the annotation of an utterance. The waveform, spectrogram, f0 and intensity curve were used (a) to determine the word boundaries and (b) to segment the disyllables into consonants and vowels. The segment boundaries were identified on the basis of the consistent application of the segmentation criteria which are described in detail in DePaolis et al. (2008) and to which the interested reader is referred to for more information. Since the authors did not specify any criteria for the segmentation of /l/, the onset and offset of the lateral approximant were determined on the basis of the discontinuity on the spectrogram in the intensity and/or frequency of the formants of /l/ and those associated with the preceding or following vowel.

Reliability Of Segmentation

The words and segments were annotated by the author IDC. Approximately 12% of the data (n = 250) was re-segmented by the author MP. The reliability of the placement of the segment boundaries was analysed by means of the Pearson's product-moment correlation between the segment durations of both annotators. The correlation between both sets of annotations was 0.99 ($p < .0001$). As annotators located the boundaries of segments at virtually the same time points, it was not deemed necessary to carry out separate reliability checks for duration, intensity and f0 of the vowels (see section below) as the Praat script would have returned extremely similar values.

Acoustic Analysis

The disyllables that were identified by the procedure described above were analysed acoustically for the prosodic cues which are relevant to the perception of syllable prominence, i.e. duration, intensity and f0. The acoustic analyses were carried out by means of a PRAAT script (Boersma & Weenink, 2014). Each of the three acoustic parameters was measured for the vowels of the disyllables only, not for the entire syllable. This was done to reduce potential effects of syllable composition on the measurements. Duration (in ms) was measured from the start to the end of each vowel. Intensity was measured in dB as the mean energy averaged over the total number of analysis frames in the vowel. F0 was determined by means of the PRAAT autocorrelation method and expressed in Hz as the mean f0 averaged over the total number of analysis frames in the vowel. Intensity and f0 were analysed by the PRAAT analysis settings adjusted to child speech, i.e. f0 range was set at 150-800 Hz and intensity range was set at 0-100 dB. It should be mentioned that intensity measurements in general need to be treated with caution. Intensity is very sensitive to background noise and recording quality. Since participants in this study were highly mobile and clip-on microphones were not used, we controlled for possible problematic intensity values by applying rigid selection criteria, cleaning the collected dataset for outliers and most importantly: by computing the ratio between the intensity measurements of the two syllables. The purpose of this study was to investigate the acoustic differentiation between syllables of utterances. Therefore, a ratio between the measurements in each syllable was computed to quantify this differentiation (i.e. durationV1/durationV2 and intensityV1/intensityV2). This also had the effect of normalising the intensity in louder utterances. The perceptual f0 distance between the first and the second vowel in each disyllable was calculated by means of the formula $|39{,}86\ log10(f0V2/f0V1)|$. This specifies the perceptual distance in semitones between the first and the second vowel in each disyllable.

The dataset was split into four subsets, i.e. words-CI, words-NH, babble-CI and babble-NH. The data in each subset were cleaned by means of the interquartile rule (IQR). Any measurement above or below the IQR threshold was identified as an outlier. The final dataset consisted of 2076 disyllables of which 925 were CI utterances and 1151 were NH utterances (for numbers per utterance type and participant, see Table 3). For the pitch distance 105 disyllables were considered outliers, for intensity ratios there were 49 outliers and for duration ratios 111 outliers.

INSERT TABLE 3 ABOUT HERE

Statistical approach

Linear mixed models (LMM) were used for the data analysis (Baayen, 2008). LMM are particularly suited to analyse longitudinal corpus data because of their hierarchical structure: the observations (n = 2076) are measured at different time points embedded in different participants (n = 18) and different groups (i.e. CI or NH). Moreover, linear mixed models are robust to missing and unequal numbers of observations for participants and time points. Importantly, when examining the effects of independent variables LMM take into account variation at the participant level as well as variation over time.

The analyses were carried out in R (R Core Team, 2013) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) to generate models for each prosodic cue. Every model consisted of random and fixed effects. In all analyses the random part provided random intercepts and slopes per individual. The fixed effects of interest were 'age' to investigate whether a cue showed development over time, 'group' (i.e. NH or CI) and ‘utterance type’ (i.e. babble, or words). The analyses described in the results section detail which fixed effects and interactions between fixed effects yielded the best fitting model. The statistical procedure consisted of two phases. A hierarchical approach was taken to build the models as random and fixed effects were added in stepwise fashion from a simpler to a more complex model. At each step, a likelihood ratio test was carried out to arrive at the best-fitting model for the data, i.e. the model explaining the largest amount of variance with the fewest predictors. In the second phase we took the best-fitting model and checked which effects were significant predictors. The estimates (henceforth E), standard errors (S.E.), t- and p-values of the fixed effects of the best-fitting models are reported in the results section.

Research question one was addressed by comparing the effect of words on the acoustic cues in the groups and considering their combined effect. For research question two, developmental slopes of cues were compared between groups. In addition, group and individual means or medians for each cue were considered in order to examine whether groups were approximating similar stress patterns.

RESULTS

Table 4 displays the means and standard deviations of the acoustic cues in babble and words in each group and at individual level.

INSERT TABLE 4 ABOUT HERE

Pitch Distance

INSERT FIGURE 2 ABOUT HERE

INSERT FIGURE 3 ABOUT HERE

For the pitch distance, the best fitting model consisted of the fixed effects of age, participant group (CI or NH), utterance type (babble or word) and the interaction between the latter two. The results are illustrated in Figures 2 and 3 and Table 1 in the supplementary material gives the output of the statistical model.

The estimate for this statistical model was 1.958 (*S.E.* = 0.186, t = 10.550, p < 0.001). The fixed effect of age evidenced development over time (E = 0.045, *S.E.* = 0.016, t = 2.734, p = 0.011), as pitch distances increased when the infants got older. The fixed effect of participant group was significant (E = 1.300, *S.E.* = 0.241, t = 5.400, p < 0.001), as the NH disyllables had bigger pitch distances compared to those from the CI group. (For descriptive statistics at the group and individual level, see table 4). Although the fixed effect of utterance type significantly improved the model, it did not reach significance (E = -0.123, *S.E.* = 0.173, t = -.709, p = 0.478).

Additionally there was a significant interaction between participant group and utterance type (E = -0.592, *S.E.* = 0.221, t = -2.679, p = 0.008), indicating a difference in how the two groups instantiated pitch distances in words and babble, as NH children had larger pitch distance increases than children with CI. At an individual level, eight out of the nine NH children increased their mean pitch distances in words (Table 4), although note that the child who did not show evidence of an increase also had the fewest data points in words (Table 3). Much smaller increases from babble to words were also seen in seven out of the nine children with CI (Table 4).

The significant interaction between participant group and utterance type was further examined through post-tests. NH infants had significantly smaller pitch distances in babble than in words (E = -0.715, *S.E.* = 0.165, z = -4.346, p < 0.001). Although the same tendency was present in the CI group, it was much reduced and was not statistically significant (E = -0.123, *S.E.* = 0.173, z = -0.709, p = 0.887). Possibly, the small difference between babble and words in the CI group may have counteracted the larger difference in the NH group and prevented the fixed effect of utterance type from reaching significance. Furthermore, comparisons of the babble of both groups showed that the NH group already had larger pitch distances at the babbling stage ($E = 0.707$, $S.E. = 0.222$, $z = 3.190$ $p = 0.007$). When comparing the words of both groups, the difference was even larger ($E = 1.300$, $S.E. = 0.241$, $z = 5.400$, $p < 0.001$).

INSERT FIGURE 4 ABOUT HERE

INSERT FIGURE 5 ABOUT HERE

If we are also interested in the direction of the stress pattern, the absolute numbers given by the semitone conversion formula are not informative. Instead, the signed numbers should be considered, as negative numbers indicate higher f0 on the first syllable and positive numbers are obtained with the opposite pattern. These numbers are represented by the boxplots in Figures 4 and 5, with the negative polarity plotted in the upper half of the x-axis for ease of reading. For the NH participants, Figure 5 shows a tendency towards the trochaic pattern at the babbling stage, where 5 children have medians in the trochaic range. This pattern becomes more pronounced in words, where 8 children have medians in the trochaic range, along with increased distances. (Participant NH6, who does not show a trochaic tendency also has few data points for words, see table 3). For children with CI, a trochaic tendency does not seem apparent at the babble stage, as only two children have medians in the trochaic range. In this group, in addition to less clear increases of pitch distances in words, there is also less of a trend towards a trochaic pattern, as only 4 children have medians in the trochaic range for words.

Intensity Ratio

INSERT FIGURE 6 ABOUT HERE

INSERT FIGURE 7 ABOUT HERE

The best-fitting model for the intensity ratio consisted of the fixed effects of age, group and utterance type (see Table 2 in the supplementary material). The intercept of this model was estimated at 1.009 (*S.E.* = 0.007, t = 140.136, p < 0.001). The effect of age significantly improved the model and approached significance (E = 0.001, *S.E.* = 0.001, t = 1.925, p = 0.060). The groups also differed in their use of this cue, with the NH group making overall slightly larger intensity differences between syllables (E = 0.023, *S.E.* = 0.009, t = 2.510, p = 0.019). The main effect of utterance type (E = -0.015, *S.E.* = *0*.005, t = -2.811, p = 0.005) suggests that intensity was not used in the same manner for babble and words, with Figures 6 and 7 confirming that the intensity ratio was smaller for babble in both groups. No post-hoc tests were carried out since the interaction between group and utterance status did not significantly improve the fit of the model. The mean ratios indicate that in words, both groups place greater intensity on the first syllable (Table 4), in accordance with a trochaic pattern. For babble, the mean ratio at group level may lead to the assumption that the CI group is qualitatively deviating from the NH pattern, as the mean ratio just below 1 suggests more intensity on the second syllable, unlike the pattern seen in the NH group. However, this conclusion is difficult to support, since individual mean ratios (Table 3) show four out of the nine children with CI have the unexpected pattern, but three of the NH children also show evidence of higher intensities on the second syllable in disyllabic babble.

Duration Ratio

INSERT FIGURE 8 ABOUT HERE

INSERT FIGURE 9 ABOUT HERE

The best fitting model for the duration ratio included age, utterance type, participant group and the interaction between utterance type and participant group (see Table 3 in the supplementary material) in the fixed effects. Figures 8 and 9 illustrate the findings. The intercept was estimated at 0.908 (*S.E.* = 0.047 $t = 19.254$, $p < 0.001$). There was a significant effect of age on the duration ratio ($E = 0.013$ *S.E.* = 0.004, $t = 3.326$, $p = 0.002$), as duration ratios increased over time. The fixed effect of participant group significantly improved the fit of the model, but did not reach significance ($E = -0.072$, *S.E.* = 0.063, $t = -1.131$, $p = 0.269$). Utterance type was a significant main effect, as lexical disyllables had larger duration ratios in both groups ($E = -0.165$, *S.E.* = 0.037, $t = -4.363$, $p < 0.001$). Post-tests on the significant interaction effect between groups and utterance type ($E = 0.155$, *S.E.* = 0.050, $t = 3.118$, $p = 0.002$) showed that only the CI group made a significantly smaller duration ratio in babble compared to words ($E = -0.165$, *S.E.* = 0.038, $z = -4.363$, $p < 0.001$). No other comparisons reached statistical significance. The ratios below 1 indicate that the second syllable was longer for both groups and in both types of disyllables (see Table 4). In words, ratios increase closer to 1. This increase is evident in all children with CI and in seven out of the nine NH children (Table 4).

DISCUSSION

The present study examined the impact of the emerging vocabulary on the ability to produce the phonetics of prosody in children with CIs and with NH. A second aim was to examine the developmental trajectories of the acoustic cues in both groups and to find out how early children with CI start to approximate the patterns seen in NH children. To this end, the pitch distance, intensity and duration ratios of babbled disyllables and first words in children with CI and NH children were compared. It was hypothesized that the CI group would be able to converge towards the same pattern as the NH group, although it was unclear how early this ability would emerge, nor how robust it would be. Because of the restrictions in signal processing, reduced use of f0 and possibly intensity by the CI group in comparison to the NH group was predicted, whereas no differences in duration use were predicted between the groups. We first discuss the results for each cue and then draw evidence from all three cues to bear on the research questions.

In both groups, mean pitch distances increased over time, but this happened to a far lesser degree in the utterances of the CI groups: the main effect of group indicated that pitch distances in NH disyllables were higher than in CI disyllables. This is likely due to processing limitations of the implant. This effect was exacerbated in words, as only the NH group significantly increased their pitch distances from babble to words. Although a similar tendency was present in the CI group, it did not reach statistical significance. Development over time was evident in both groups, however, when comparing the regression lines in figures 2 and 3, the most dramatic increase seems to be occurring in the words of the NH group. Figure 5 suggests that this increase in pitch differentiation was accompanied by a shift towards a more clearly trochaic pattern in disyllabic words. In essence, a trochaic tendency at the babble stage appears to become crystallised at the word level for the NH group. Figure 4 does not appear to give evidence of a trochaic pattern for the CI group at babble stage; at word level, there is some evidence of a shift of pitch distances into more trochaic values in four participants, nevertheless the move towards trochaic values is far less evident in this group. Therefore, in answer to research question one, no enhancement for f0 use was seen for the words of the CI group. For research question two, the effect of the implant's processing limitations were visible, in that overall development was slower and smaller in the CI group and there was no evident convergence towards the ambient trochaic pattern.

For intensity, little overall development over time was seen in ratios in both groups, but an enhancement of intensity differences was present in words in comparison to babble in both groups. The fact that including an interaction between group and utterance type did not increase the fit of the model indicates that both groups did this to the same degree (see also Table 4). Nevertheless it is unclear whether the two groups truly follow the same developmental trajectories. On group means (Table 4), the CI group appears to start out from a pattern in babble which deviates from the NH pattern, as the ratio below 1 suggests that intensity was higher on the second syllable, contra the predominant trochaic (strong-weak) pattern for Dutch. Recall however that four out of the nine children with CI displayed the unexpected pattern, but three of the NH children also showed evidence of higher intensities on the second syllable in babble. Judgement on whether the developmental trajectory of the CI group represents an atypical pattern of intensity use should therefore be withheld. In words, all NH children and eight Children with CI transitioned to ratios above one. It was again the case that the CI group had smaller ratios than the NH group, i.e. smaller differences between syllables in terms of intensity. The answer to the first research question is entwined with the answer to the second one: since neither group made very clear use of intensity both in terms of size of difference nor in terms of the direction of the stress pattern, it is difficult to tell whether children with CI truly expand their intensity ratio on words. Nominally, the implant seems to have little effect on intensity, but note that there is also very little development over time in both groups, therefore it is not clear how functional the use of intensity is in either group.

For duration, the mean group ratios (Table 4) below one in babble and words indicate that the second syllable was longer for both groups. Since these were disyllables spoken in isolation, it is very likely that utterance-final lengthening (an increase in duration of the final syllable) is at work and may potentially obscure durational effects of prosodic prominence (White, 2014). Interestingly, the duration ratios in words are closer to 1, which suggests that the difference between syllables lessens, giving a less modulated pattern. In terms of individual data (Table 4), all the children with CI increased their mean duration ratios in words, but in the NH group, two slightly decreased their mean ratios from babble to words. In response to the first research question, the words of children with CI are in fact less modulated, but it is unclear whether this is due to a failure to produce word stress, or differences in final lengthening in the groups. For inferential statistics, age, utterance type and group were needed for the best fit of the model with the first two reaching significance. When considering the regression lines in Figures 8 and 9, a developmental trend appears to be present in the babble and words of the CI group, whereas in the NH group only words show an increase in duration ratios. Statistically, this was reflected in a significant interaction between utterance type and group, followed up by post-tests which showed that only the CI group significantly increased their ratios from babble to words. In terms of research question two, the difference between groups ran contra to our prediction for duration, since this cue is available to CI listeners. However, as durational phenomena in this dataset are unlikely to be a simple result of prosodic prominence at the level of the word, it is difficult to make strong statements on the use of duration for signalling prosody in children with CI. It is striking that although both groups had duration ratios in words which indicated that syllables were less modulated, this was only statistically significant for the CI group.

Why would words become less modulated for children with CI? It has been suggested that children may simplify rhythmic properties to give equal weight to each syllable in an utterance when acquiring new linguistic structures (Snow, 1994; Redford & Sirsa, 2011), as a more isochronous rhythm is thought to be easier to acquire (Goffman, 1999; Payne, Post, Astruc, Prieto, & Vanrell, 2011). The children with CI started producing canonical babble and words later, and therefore had less opportunity to practice speech planning and articulation than the NH group. It may be that the effort of integrating a clear adult target and attempting to produce it is more costly to the less mature speech planning system of the CI group. Indirect evidence for lower articulatory maturity in this group comes from Vanormelingen, De Maeyer & Gillis (in press), who found lower articulation rates for the infants with CI included in the present study than NH age matches. Articulation rate has been used as a proxy for speech maturity as children's articulation rate slowly increases towards adult values over the first ten years of life (Lee, Potamianos, & Narayanan, 1999; Redford, 2014). Similarly, the infants with CI included in this study were also shown to have significantly less complex syllable structures in words than both age and vocabulary matched NH controls (van den Berg, 2012), and Faes et al. (2015) showed that their word accuracy was significantly more affected by phonolgical complexity than NH controls. Therefore, the rhythmic simplification in words may fit with a general tendency for simplified linguistic structure in the CI children’s speech.

Summarising the findings for the three cues, and in response to the main research question, the advent of recognizable word use did not trigger the same expansion of prosodic differentiation in children with CI as in NH children; this was most clearly visible on f0. For intensity, although participants with CI had smaller ratios, they did increase the difference between the first and second syllable in words. However, it was unclear to what degree this cue was used reliably to signal a trochaic pattern by either group. Duration ratios indicated that for the children with CI, the transition to words may have posed an additional articulatory challenge. For individual data, only descriptive statistics were considered and these indicated larger variation in the CI group for f0 and intensity in terms of the direction of stress; conversely, duration showed less variation in the CI group than the NH group.

With respect to the secondary research question, which concerned developmental trajectories in groups, the results may be indicative of a developmental lag: possibly children with CI only reach the level of prosodic differentiation in words NH children already display when they are babbling, as the group results for pitch distance and intensity approached more trochaic levels in words (although individual variation should be kept in mind). If the initial auditory deprivation is taken into account, the idea of a lag is inherently appealing: after all, children with CI have had less aural exposure to language than the comparison group, and they should therefore still be catching up when they are in the word stage. The lack of auditory stimulation may become compounded by atypicalities in babbling: Infants with CI have had less time to explore their own speech production via auditory feedback, and work by Koopmans-van Beinum, Clement and Den Dikkenberg-Pot (2001) as well as Schauwers, Gillis and Govaerts (2008) has reported less variegated babble in infants with CI. For speech production, it may be that babbling provides an important training ground for the phonetic features of the ambient language, so that these can quickly become stabilized once vocabulary items appear. Therefore, the initial lag in babbling development may have contributed to difficulties in starting to approximate the native stress pattern at word level.

In order to strengthen the conclusions of the present study, perceptual ratings of prosodic prominence in children's utterances will be needed. Considering cues separately does not give a complete picture of prosodic abilities, since prosody is a multi-dimensional phenomenon and cues may interact and be in trade-off relations (Lieberman, 1960). This would clarify the somewhat ambiguous findings for intensity. In addition, we have only presented descriptive statistics for the directions of stress patterns, and future investigations should contain inferential statistics. Furthermore, since the analyses were carried out on spontaneous recordings, two of the participants have few data points: one infant with CI at the babbling stage, and one NH infant at the word stage. When comparing two groups of 9 infants each, this is likely to affect the robustness of the findings, although the statistical treatment is designed to take account of differential datapoints per individuals and mitigate this. Finally, more controlled recordings in sound-insulated laboratories and the use of clip-on microphones would be particularly beneficial for investigations of intensity use.

Keeping these reservations in mind, it remains interesting to consider the lack of top-down effects on prosodic modulation in the words of children with CI in terms of the clinical implications and future research. The fact that phonetic development does not seem boosted to the same degree in words raises the question of the nature of the developmental lag: is this merely slow phonetic development, brought about by a lack of auditory stimulation and verbal practice during the babbling stage, or does this represent a more deep-seated problem in abstracting phonological patterns from the statistical regularities of the ambient language, as McKean, Letts & Howard (2013) found to be the case for NH children with language delay? Houston and Bergeson (2014) suggested that in infants with CI, it is not just the perception of speech which is affected, but attention to speech may also be atypical. The causal relations between attenuated perception and attention are still under investigation, but one conclusion from this work and the present study is that interventions should strive to highlight linguistic structure in words, since pattern extraction and attentional mechanisms may be sub-optimal in this population.

Another area for further investigation concerns the relation between the present results and later difficulties with speech intelligibility in children with CI (Flipsen, 2008; Montag, AuBuchon, Pisoni, Kronenberger, 2014). Lenden and Flipsen (2007) reported problems with stress production in conversational speech samples of children with CI, and Hide (2013) found weaker phonetic cue use even in correctly stressed nonsense words. Can this profile in older children with CI be related back to their early speech development? Is it the case that children who have more clearly trochaic items, either in babble or at the transition to words, also go on to develop more fluent and intelligible speech? The literature review of typical language development has shown that the transition to words may be particularly important, since vocabulary development acts as a kind of bootstrap for phonetic development. In terms of the present research, we could ask whether those four children with CI whose use of f0 in words showed evidence of an emerging trochaic pattern also have more advanced abilities in the kinds of syllable structures they can produce and in terms of their intelligibility, at that particular time-point and in further development. In order to tackle these questions, research will need to determine which level of analysis is the most informative: is it enough to take children's ability to use individual cues as an indicator of competence? It may be that perceptual ratings of stress have stronger predictive value for intelligibility, as they encompass all acoustic aspects of prosody (Lieberman, 1960). Such research is likely to need larger samples and necessitate more complex modelling of developmental trajectories (e.g. Ullman, 2001) than presently available. However, we hope to have drawn the research community's attention to the transition from babble to first words as a potentially promising area of investigation in terms of early intervention and prediction.

In conclusion, we carried out a phonetic production study of prosody in disyllabic babble and first words of CI and NH infants. The results indicated that infants with CI had weaker acoustic cue use, broadly in line with the signal processing limitations of the implant. In addition, infants with CI did not benefit to the same degree from lexical effects, as they did not show the same enhancement of phonetic cues on first words as NH infants did.

References

Baayen, H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge: Cambridge University Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Bates, E., & Goodman, J. (1997). On the inseparability of grammar and the lexicon: evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes, 12*, 507-584.

Beckman, M. E. & Edwards, J. (2000). The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development*, *71*(1), 240–249. doi:10.1111/1467-8624.00139

Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer. 5.4. Retrieved 5 October, 2014, from http://www.praat.org.

Carter, A. K., Dillon, C. M., & Pisoni, D. B. (2002). Imitation of nonwords by hearing impaired children with cochlear implants: suprasegmental analyses. *Clinical Linguistics & Phonetics*, *16*(8), 619–638.

Chapman, K. L., Hardin-Jones, M., Schulte, J., & Halter, K. A. (2001). Vocal development of 9-month-old babies with cleft palate. *Journal of Speech, Language, and Hearing Research*, 44(6), 1268-1283.

Curtin, S., Campbell, J., & Hufnagle, D. (2012). Mapping novel labels to actions: how the rhythm of words guides infants' learning. *Journal of Experimental Child Psychology*, *112*(2), 127–40. doi:10.1016/j.jecp.2012.02.007

Cutler, A. (2005). Lexical stress. In D. Pisoni & R. Remez (Eds.), *The handbook of speech perception* (pp. 264-289). Oxford: Blackwell.

Daelemans, W., Gillis, S., & Durieux, G. (1994). The acquisition of stress: a data-oriented approach. *Computational Linguistics, 20*(3), 421-451.

De Clerck, I., Pettinato, M., Verhoeven, J. & Gillis, S. (in press). Is prosodic production driven by lexical development? Longitudinal evidence from babble and words. *Journal of Child Language*.

Demuth, K. (1996). The prosodic structure of early words. In J. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 171-184). Mahwah, N.J.: Lawrence Erlbaum Associates.

DePaolis, R. a., Vihman, M. M., & Kunnari, S. (2008). Prosody in production at the onset of word use: a cross-linguistic study. *Journal of Phonetics*, *36*(2), 406–422. doi:10.1016/j.wocn.2008.01.003

Dinnsen, D. A., Green, C. R., Morrisette, M. L., & Gierut, J. A. (2011). On the interaction of velar fronting and labial harmony, *Clinical Linguistics & Phonetics*, *25*(3), 231–51. doi:10.3109/02699206.2010.522300

Flipsen, P., Jr. (2008). Intelligibility of spontaneous conversational speech produced by children with cochlear implants: a review. *International Journal of Pediatric Otorhinolaryngology*, 72(5), 559-564.

Friederici, A., Friedrich, M., & Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Current Biology, 17*, 1208-1211.

Gerken, L. (1994). A metrical template account of children's weak syllable omissions from multisyllabic words. *Journal of Child Language, 21*(3), 565-584.

Green, T., Faulkner, A., & Rosen, S. (2004). Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *The Journal of the Acoustical Society of America*, *116*(4), 2298-2310.

Goffman, L. (1999). Prosodic influences on speech production in children with specific language impairment and speech deficits. *Journal of Speech, Language, and Hearing Research*, *42*(6), 1499. doi:10.1044/jslhr.4206.1499

Heisler, L., Goffman, L., & Younger, B. (2010). Lexical and articulatory interactions in children's language production. *Developmental Science*, *13*(5), 722–730. doi:10.1111/j.1467-7687.2009.00930.x

Holt, C. M., & McDermott, H. J. (2013). Discrimination of intonation contours by adolescents with cochlear implants. *International Journal of Audiology*, *52*(12), 808–15. doi:10.3109/14992027.2013.832416

Holt, C., & Fletcher, J. (2015). Perception and interpretation of low-onset rising tunes by prelingually deaf cochlear implant users. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow: The University of Glasgow.

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 : A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *15*(2), 136–146.

Houston, D. M., & Bergeson, T. R. (2014). Hearing versus listening: attention to speech and its role in language acquisition in deaf infants with cochlear implants. *Lingua*, 139, 10-25.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginning of word segmentation in English-learning infants. *Cognitive Psychology*, 39, 159–207.

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience*, 16(8), 1412-1425.

Koopmans - van Beinum, F. J., Clement, C. J., & van den Dikkenberg-Pot, I. (2001). Babbling and the lack of auditory speech perception: a matter of coordination?

Developmental Science, 4(1), 61-70.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, *105*(3), 1455-1468. doi:10.1121/1.426686

Lenden, J. M., & Flipsen, P. (2007). Prosody and voice characteristics of children with cochlear implants. *Journal of Communication Disorders*, *40*(1), 66–81.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America, 32*(4), 451-454.

Lieberman, P. (1984). *The Biology and Evolution of Language.* Cambridge: Harvard University Press.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Mahwah: Lawrence Erlbaum.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), 101–111. doi:10.1016/S0010-0277(01)00157-3

McKean, C., Letts, C., & Howard, D. (2013). Developmental Change Is Key to Understanding Primary Language Impairment: The Case of Phonotactic Probability and Nonword Repetition. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1579–1594. doi:10.1044/1092-4388(2013/12-0066)

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), 459-467.

Molemans, I., van den Berg, R., van Severen, L., & Gillis, S. (2012). How to measure the onset of babbling reliably? *Journal of Child Language*, *39*(3), 523–52. doi:10.1017/S0305000911000171

Montag, J. L., AuBuchon, A. M., Pisoni, D. B., & Kronenberger, W. G. (2014). Speech intelligibility in deaf children after longterm cochlear implant use. *Journal of Speech, Language, and Hearing Research, 57*(6), 2332-2343.

Moore, B. C. J. (2003). Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants. *Otology & Neurotology 24*(2), 243–54.

Morgan, J. L., & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah: Lawrence Erlbaum.

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. doi:10.1093/deafed/enm012

Nakata, T., Trehub, S. E., & Kanda, Y. (2012). Effect of cochlear implants on children's perception and production of speech prosody. *The Journal of the Acoustical Society of America, 131*(2), 1307–1314.

Nicholson, H., Munson, B., Reidy, P. F., & Edwards, J. R. (2015). Effects of age and vocabulary size on production accuracy and acoustic differentiation of young children ' s sibilant fricatives. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow: The University of Glasgow.

O ' Halpin, R. (2010) . The perception and production of stress and intonation by children with cochlear implants . (unpublished Doctoral thesis). University College London, London). http://eprints.ucl.ac.uk/20406/

Payne, E., Post, B., Astruc, L., Prieto, P., & del Mar Vanrell, M. (2011). Measuring child rhythm. *Language and Speech, 55*, 203-229.

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), 336–351. doi:10.1097/AUD.0b013e318168d94d

Peters, R. W., Moore, B. C., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *The Journal of the Acoustical Society of America*, *103*(1), 577–587.

Pettinato, M., De Clerck, I., Verhoeven, J., & Gillis, S. (2015). The production of word stress in babbles and early words: a comparison between normally hearing infants and infants with cochlear implants. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow: The University of Glasgow.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*, 115-154.

Redford, M. A. (2014). The perceived clarity of children's speech varies as a function of their default articulation rate. *The Journal of the Acoustical Society of America*, *135*(5), 2952–2963. doi:10.1121/1.4869820

Sirsa, H., & Redford, M. A. (2011). Towards understanding the protracted acquisition of English rhythm. In Lee, W.-S., Zee, E (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (p. 1862). Hong Kong: City University of Hong Kong

Schauwers, K., Gillis, S., & Govaerts, P. J. (2008). The characteristics of prelexical babbling after cochlear implantation between 5 and 20 months of age. *Ear and Hearing*, *29*(4), 627-637.

Segal, O., Houston, D., & Kishon-Rabin, L. (2015). Auditory discrimination of lexical stress patterns in hearing-impaired infants with cochlear implants compared with normal hearing: influence of acoustic cues and listening experience to the ambient language. *Ear and Hearing*, *37*(2), 225-234 doi:10.1097/AUD.0000000000000243

Sharma, A., Dorman, M. F., & Kral, A. (2005). The influence of a sensitive period on central auditory development in children with unilateral and bilateral cochlear implants. *Hearing Research*, *203*(1), 134-143.

Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech and Hearing Research, 37*(4), 831-840.

Spencer, L. J., & Tomblin, J. B. (2009). Evaluating phonological processing skills in children with prelingual deafness who use cochlear implants. *Journal of Deaf Studies and Deaf Education*, *14*(1), 1–21. doi:10.1093/deafed/enn013

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language, 38*(1), 1-34.

Titterington, J., Henry, A., Krämer, M., Toner, J. G., & Stevenson, M. (2006). An investigation of weak syllable processing in deaf children with cochlear implants. *Clinical Linguistics & Phonetics*, 20(4), 249-269.

Torppa, R., Faulkner, A., Huotilainen, M., Järvikivi, J., Lipsanen, J., Laasonen, M., & Vainio, M. (2014). The perception of prosody and associated auditory cues in early-implanted children: the role of auditory working memory and musical activities. *International Journal of Audiology*, *53*(3), 182–91. doi:10.3109/14992027.2013.872302

Ullman, J.B. (2001). Structural equation modeling. In B. G. Tabachnich & L. S. Fidell (Eds.), *Using multivariate statistics* (pp. 653-771). New York: Allyn & Bacon.

VanDam, M., Ide-Helvie, D., & Moeller, M. P. (2011). Point vowel duration in children with hearing aids and cochlear implants at 4 and 5 years of age. *Clinical Linguistics & Phonetics*, *25*(8), 689–704. doi:10.3109/02699206.2011.552158

Van den Berg, R. (2012). *Syllables inside out. A longitudinal study of the development of* *syllable types in toddlers acquiring Dutch: A comparison between hearing impaired children with a cochlear implant and normally hearing children.* (Unpublished Doctoral thesis), University of Antwerp, Antwerp.

Vanormelingen, L., De Maeyer, S. & Gillis, S. (in press). A comparison of maternal and child language in normally hearing and children with cochlear implants, *Language, Interaction & Acquisition.*

Verhoeven, J. (2005). Belgian Standard Dutch. *Journal of the International Phonetic Association, 35*, 243-247.

Vihman, M., DePaolis, R., & Davis, B. (1998). Is there a "trochaic bias" in early word learning? Evidence from infant production in English and French. *Child Development, 69*, 935-949.

Vihman, M., & McCune, L. (1994). When is a word a word? *Journal of Child Language, 21*(3), 517-542.

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, *46*(3), 233–51. doi:10.1002/dev.20060

White, L. (2014). Communicative function and prosodic form in speech timing. *Speech Communication, 63*, 38-54.

Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, *113*(2), 234–43. doi:10.1016/j.cognition.2009.08.010

Zink, I., & Lejaegere, M. (2001). *N-CDIs: Lijsten voor communicanieve ontwikkeling. Aanpassing en hernormering van de MacArthur CDI's van Fenson et al.* Leuven: ACCO.

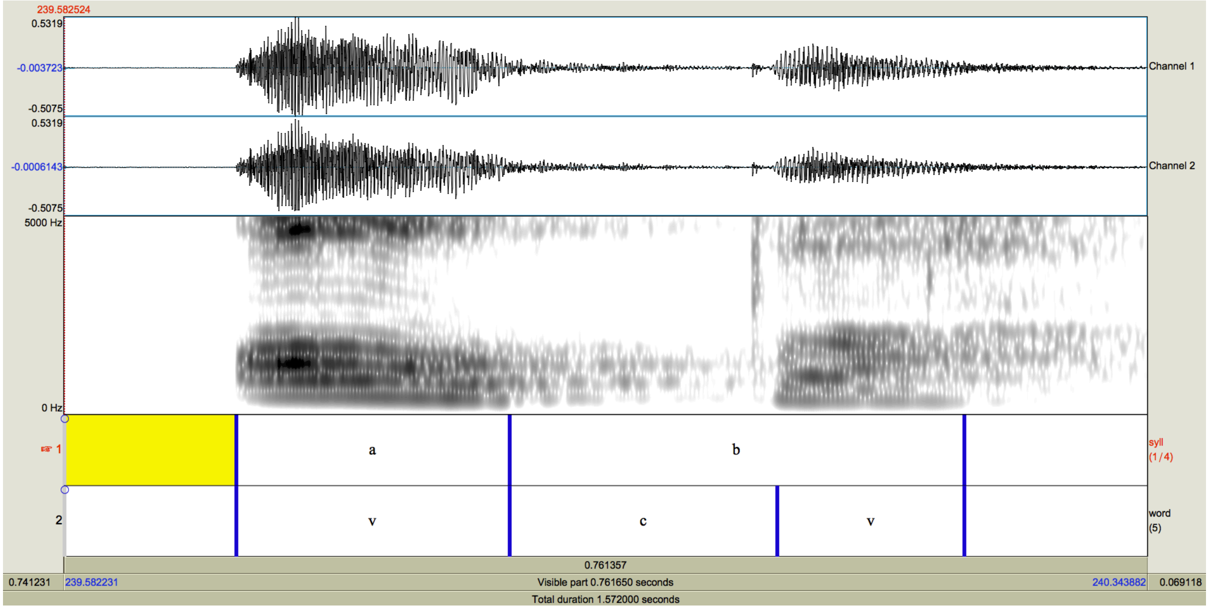

Figure 1: An annotated lexical disyllable (“auto”, /ˈautɔ/, English: “car”). Legend: a = first lexical syllable; b = second lexical syllable; v = vowel; c = consonant

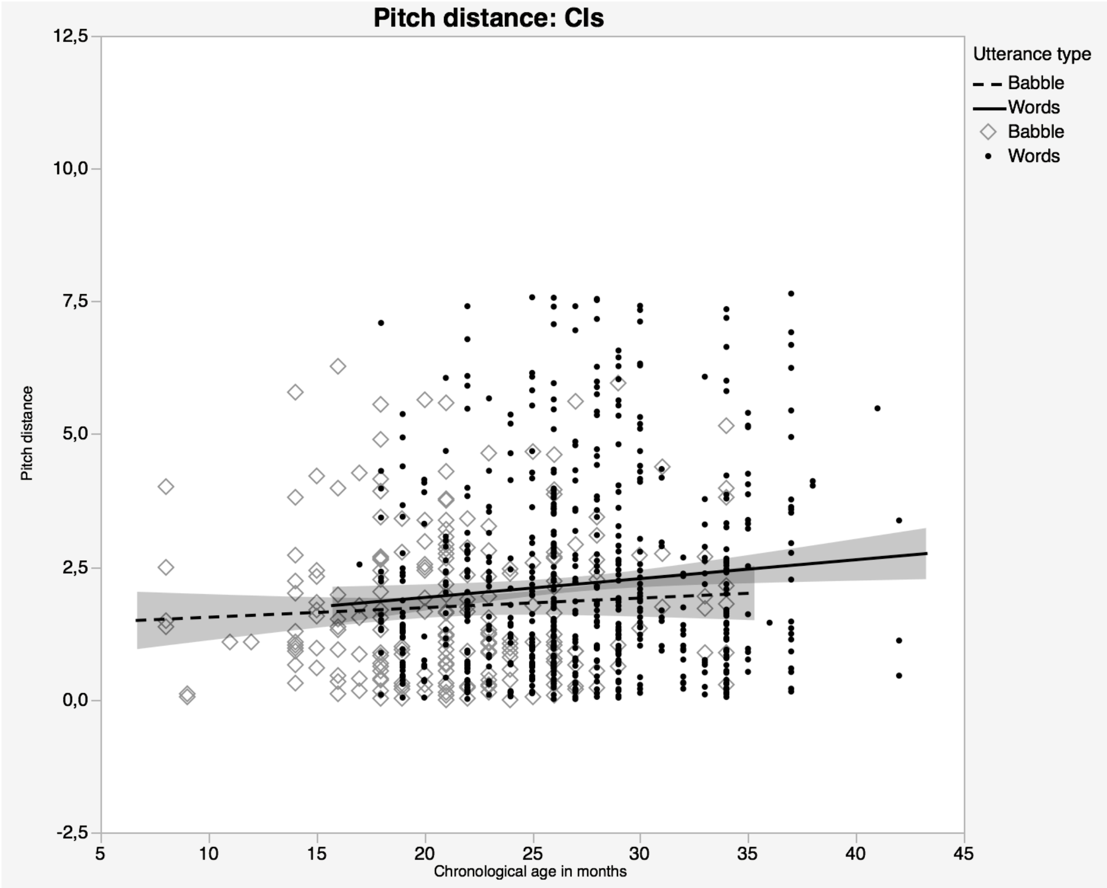

Figure 2: Scatterplots for the absolute values of the pitch distances of the children with CI. Shaded area = confidence interval

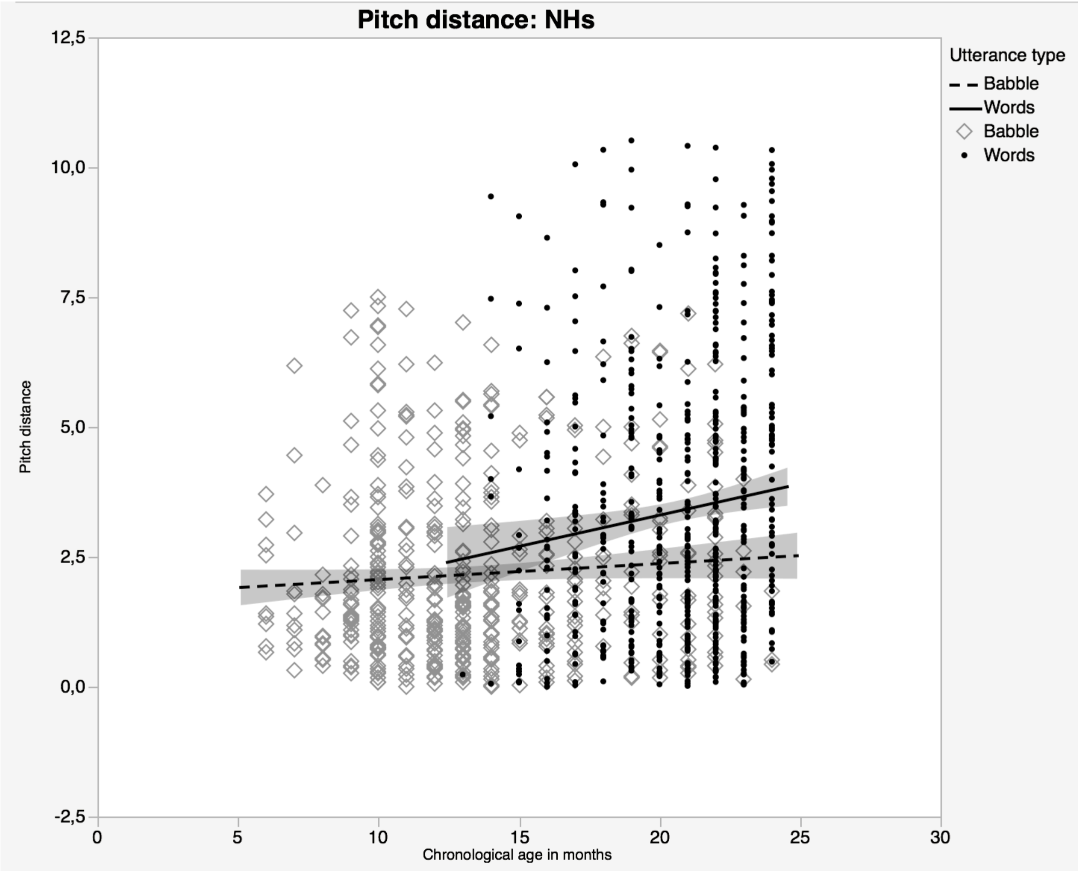

Figure 3: Scatterplots for the absolute values of the pitch distances of the NH children.

Shaded area = confidence intervals

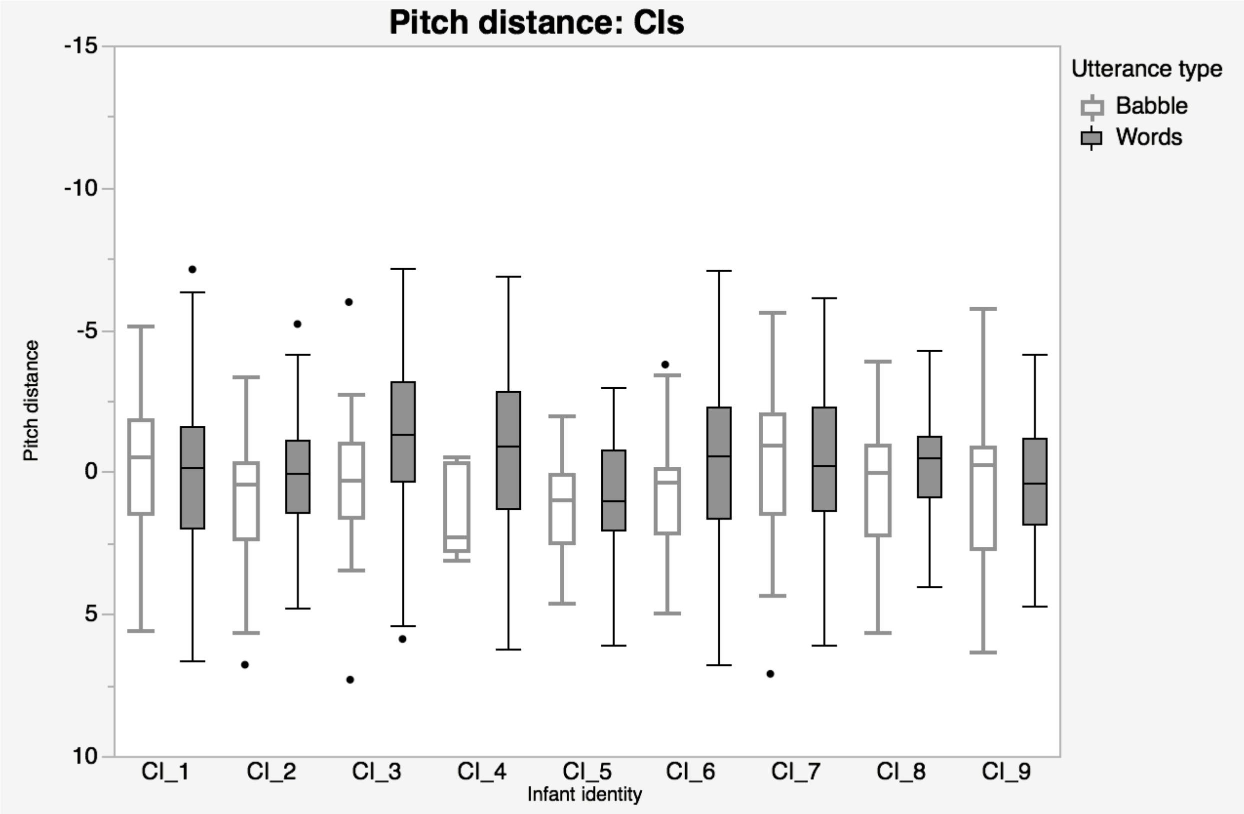

Figure 4: Boxplots for the pitch distances of babbled and lexical utterances of children with CI. Negative values indicate higher f0 on the first syllable, positive values higher f0 on the second syllable

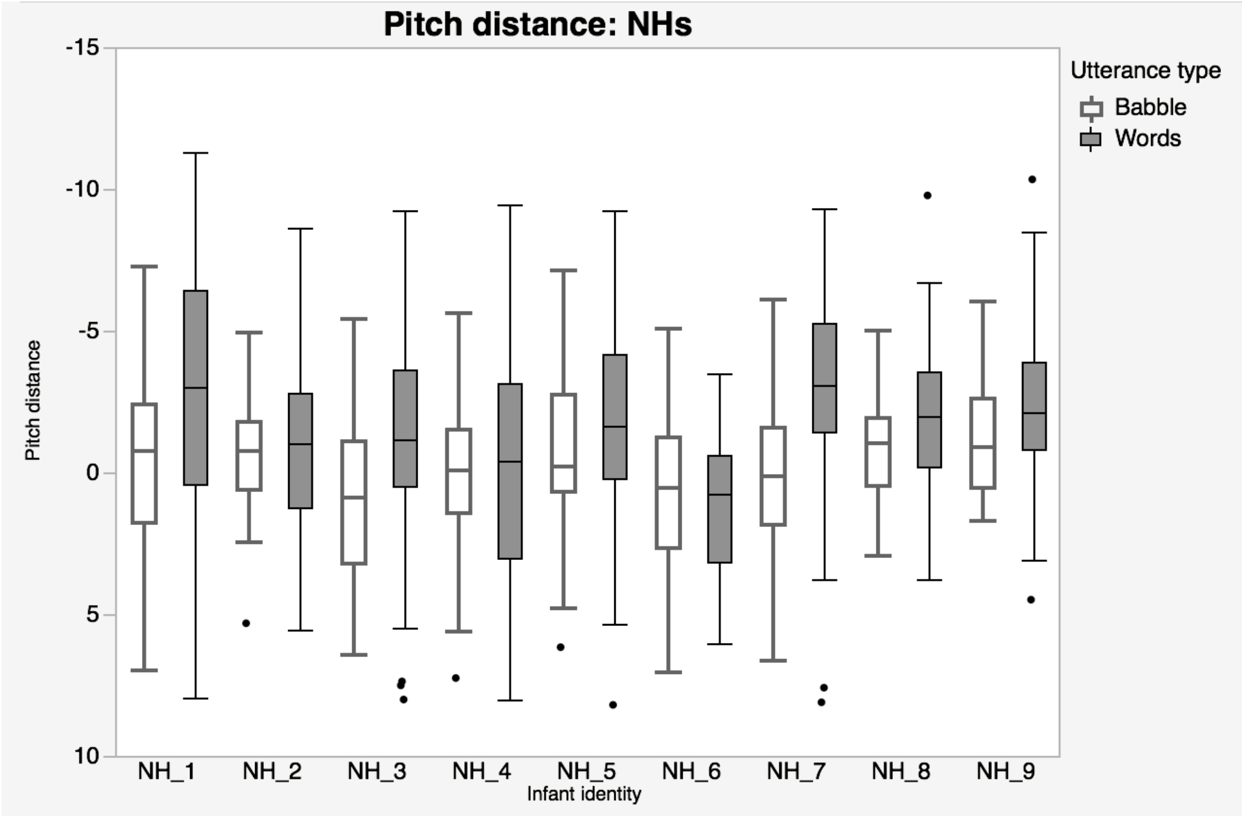

Figure 5: Boxplots for the pitch distances of babbled and lexical utterances of NH children. Negative values indicate higher f0 on the first syllable, positive values higher f0 on the second syllable

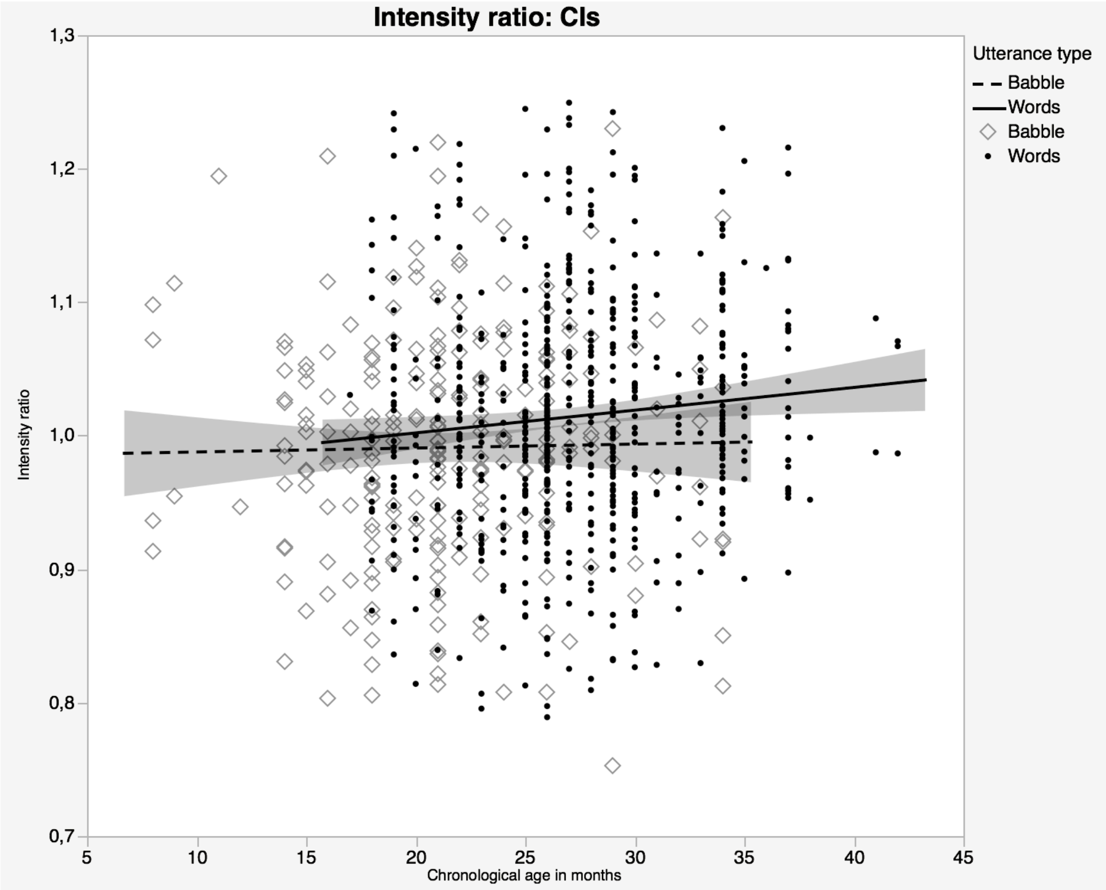

Figure 6: Scatterplots of the values of the intensity ratios for the children with CI. Shaded area = confidence intervals

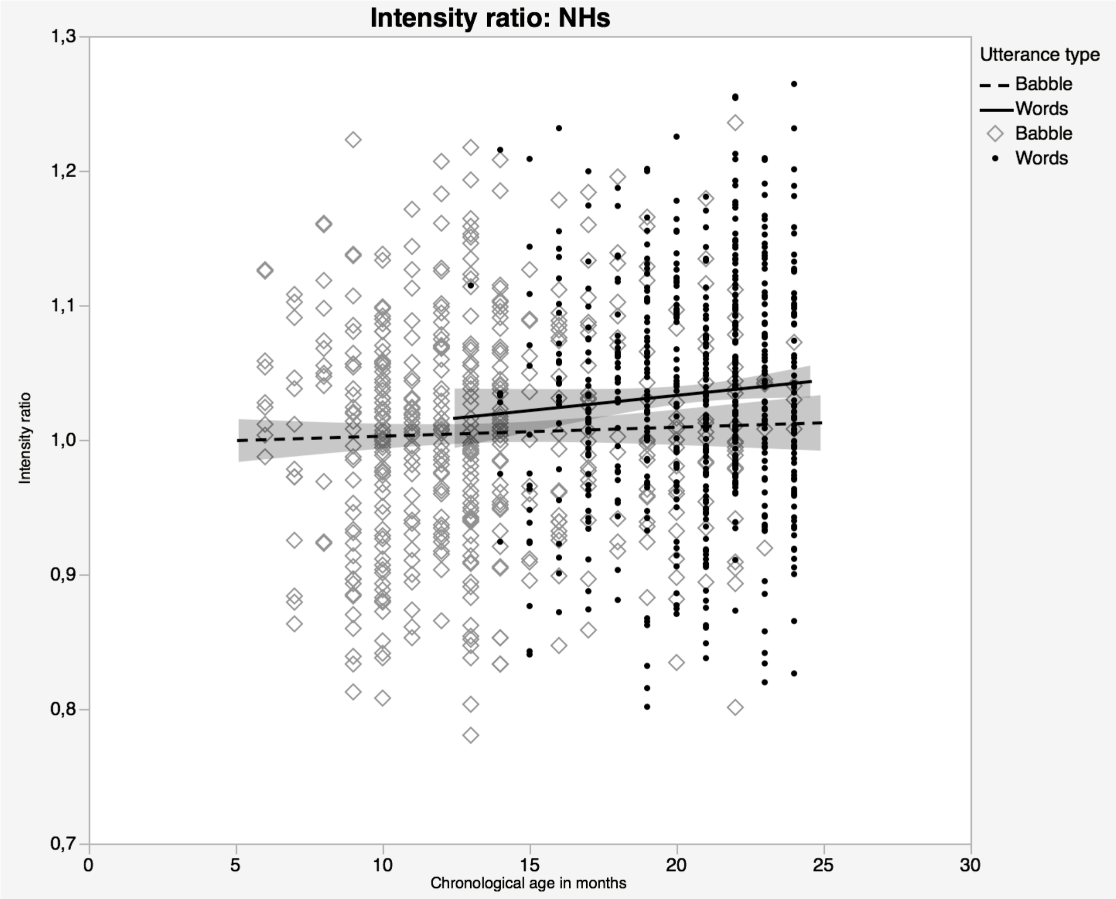

Figure 7: Scatterplots of the intensity ratios for the NH children. Shaded area = confidence intervals

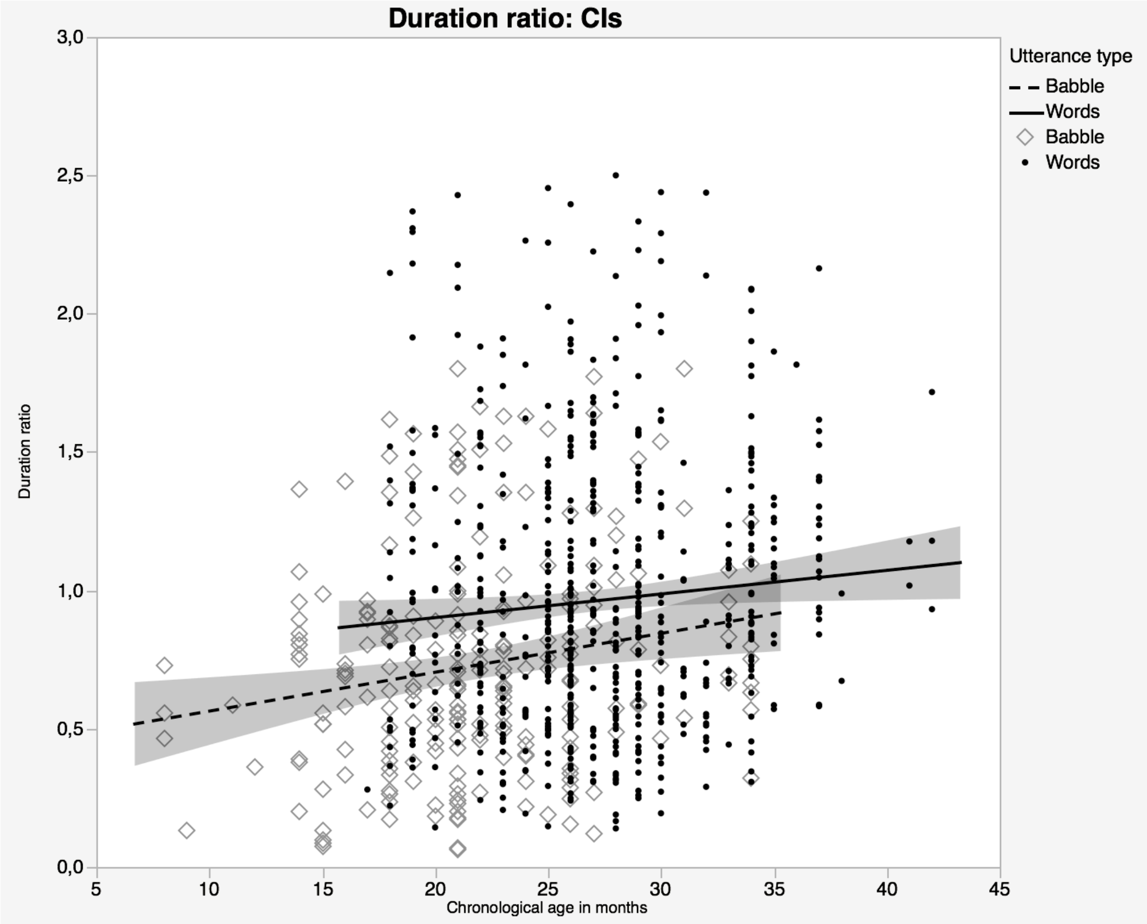

Figure 8: Scatterplots of the duration ratios for the children with CI. Shaded area = confidence intervals

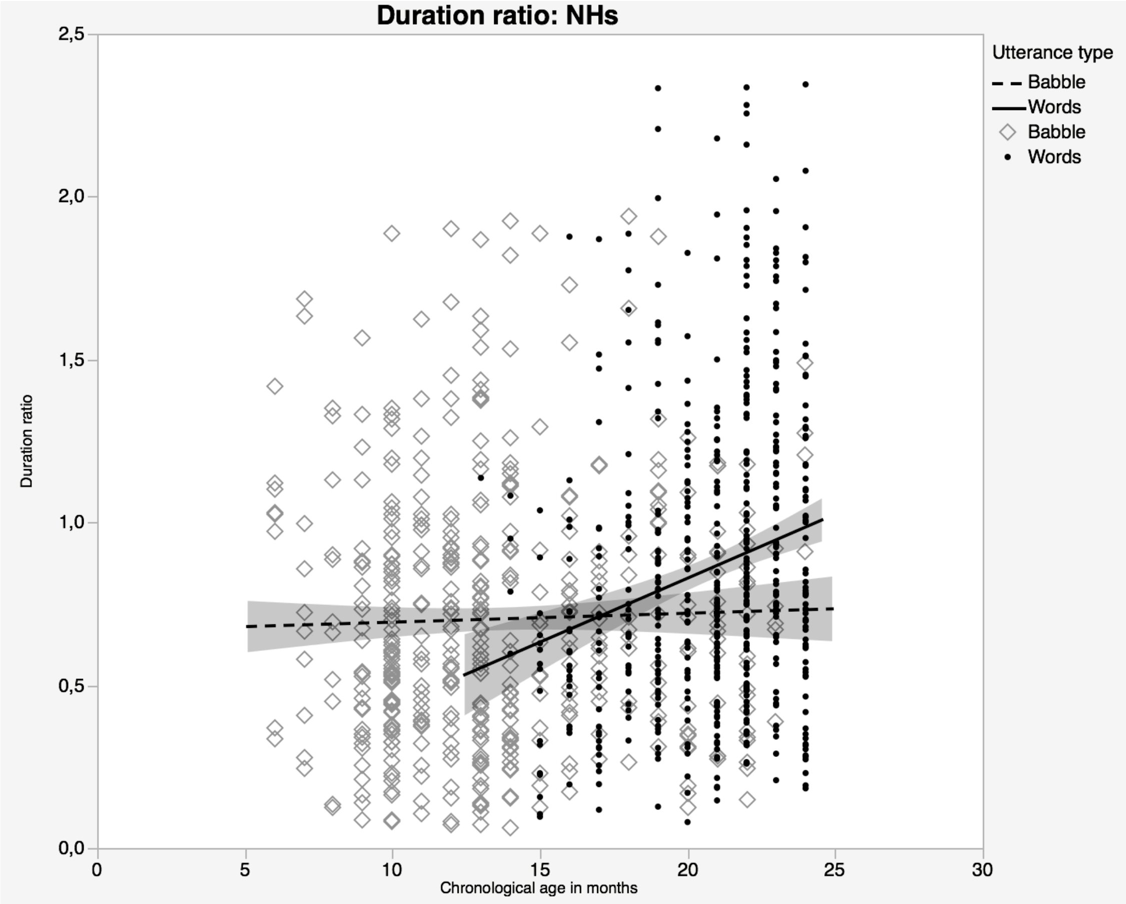

Figure 9: Scatterplots of the duration ratios for the NH children. Shaded area = confidence intervals

ID	PTA unaided (dBHL)	Age at hearing aid	PTA aided (dBHL)	Age CI	Age CI fitting	PTA CI (dBHL)
CI-1	120	0;09.03	120	1;01.05	1;2.27	47
CI-2	120	0;01.04	120	0;06.21	0;7.20	30
CI-3	115	0;01.18	113	0;10.0	0;11.20	33
CI-4	113	0;10.0	117	1;06.05	1;07.09	42
CI-5	93	0;04.24	47	1;04.27	1;05.27	35
CI-6	120	0;01.21	107	0;08.23	0;09.20	43
CI-7	117	0;04.0	107	0;05.05	0;06.04	43
CI-8	112	0;02.0	58	1;07.14	1;09.04	52
CI-9	103	0;05.08	63	0;8.21	0;9.21	32

Table 1: Auditory characteristics of the CI children in the corpus. PTA = Pure Tone Average; dBHL = decibel hearing level; HA = Hearing Aid; CI = Cochlear Implant.

ID	Gender	Age start (y;mm.dd)	Age end (y;mm.dd)	# recordings	Onset word use
CI-1	F	1;03.21	2;09.27	15	1;07.26
CI-2	F	1;03.27	2;04.27	14	1;04.28
CI-3	F	1;03.16	2;06.09	15	1;07.25
CI-4	M	1;10.23	3;06.16	14	1;08.13
CI-5	M	1;06.13	2;05.27	9	1;06.13
CI-6	M	1;03.26	2;04.26	13	1;03.26
CI-7	F	0;08.01	2;04.16	20	1;03.11
CI-8	F	1;08.21	2;10.13	13	1;11.23
CI-9	F	1;00.27	1;09.20	10	1;02.22
NH-1	M	0;06.03	2;00.01	19	1;02.09
NH-2	M	0;06.29	2;00.02	18	1;00.01
NH-3	M	0;06.00	1;11.05	19	1;01.31
NH-4	M	0;05.27	1;11.30	19	1;00.30
NH-5	F	0;06.03	1;11.29	19	0;11.00
NH-6	F	0;06.05	2;00.04	19	1;03.30
NH-7	M	0;07.02	1;11.27	18	1;04.02
NH-8	F	0;06.04	2;00.04	19	1;00.05
NH-9	F	0;08.02	1;09.28	15	1;03.00

Table 2: Ages of the individual children at the time of recording. Legend: Age start = the onset of babbling; Age end = cumulative vocabulary of at least 200 words.

	Babble	Words	Total
CI-1	40	105	145
CI-2	35	79	114
CI-3	34	149	183
CI-4	5	43	48
CI-5	24	38	62
CI-6	30	69	99
CI-7	36	87	123
CI-8	16	62	78
CI-9	20	53	73
CI_Total	**240**	**685**	**925**
CI_Mean	**26.67**	**76.11**	**102.78**
CI_SD	**10.72**	**32.67**	**43.20**
NH-1	122	161	283
NH-2	40	59	99
NH-3	47	96	143
NH-4	87	88	175
NH-5	49	45	94
NH-6	80	14	94
NH-7	36	50	86
NH-8	33	43	76
NH-9	31	70	101
NH_Total	**525**	**626**	**1151**
NH_Mean	**58.33**	**69.56**	**127.89**
NH_SD	**29.41**	**39.85**	**65.96**
Total	**765**	**1311**	**2076**
Mean	**42.50**	**72.83**	**115.33**
SD	**28.00**	**37.64**	**55.61**

Table 3: Number of selected disyllables per group (CI and NH) and per utterance type (Babble and Word). SD = Standard deviation

	Babble			Words		
ID	**Pitch distance (semitones)**	**Intensity ratio**	**Duration ratio**	**Pitch distance (semitones)**	**Intensity ratio**	**Duration ratio**
CI-1	2 (1.46)	0.97 (0.07)	0.71 (0.37)	2.11 (1.61)	1.03 (0.08)	1.02 (0.41)
CI-2	1.53 (1.6)	0.96 (0.09)	0.57 (0.31)	1.61 (1.4)	0.97 (0.09)	0.97 (0.51)
CI-3	1.49 (1.26)	1.01 (0.09)	0.78 (0.41)	2.71 (2.07)	1.01 (0.09)	0.83 (0.46)
CI-4	1.7 (1.25)	1.04 (0.07)	0.8 (0.36)	2.85 (2.13)	1.04 (0.07)	1.15 (0.34)
CI-5	1.45 (1.18)	0.99 (0.07)	0.88 (0.44)	1.93 (1.68)	1.02 (0.08)	0.95 (0.46)
CI-6	1.79 (1.5)	1 (0.1)	0.9 (0.47)	2.25 (1.79)	1.04 (0.08)	1.19 (0.61)
CI-7	1.93 (1.27)	1.01 (0.07)	0.58 (0.36)	2.27 (2.01)	1 (0.1)	0.89 (0.45)
CI-8	1.8 (1.42)	0.98 (0.1)	0.71 (0.42)	1.61 (1.51)	1.02 (0.09)	0.92 (0.55)
CI-9	1.98 (1.64)	1 (0.09)	0.74 (0.34)	1.75 (1.34)	1.01 (0.1)	0.98 (0.58)
CI_Group	1.76 (1.41)	0.99 (0.09)	0.73 (0.4)	2.18 (1.82)	1.01 (0.09)	0.96 (0.49)
NH-1	2.56 (1.91)	1.01 (0.07)	0.74 (0.4)	4.17 (2.87)	1.05 (0.09)	0.98 (0.51)
NH-2	1.61 (1.31)	1 (0.08)	0.83 (0.43)	2.47 (1.84)	1 (0.07)	0.8 (0.41)
NH-3	2.43 (1.89)	1 (0.09)	0.6 (0.34)	2.93 (2.49)	1.02 (0.08)	0.82 (0.42)
NH-4	1.95 (1.61)	1.01 (0.08)	0.68 (0.36)	3.79 (2.54)	1.01 (0.08)	0.78 (0.39)
NH-5	2.31 (1.92)	0.99 (0.08)	0.78 (0.41)	3.29 (2.47)	1.02 (0.07)	0.69 (0.37)
NH-6	2.24 (1.6)	0.98 (0.08)	0.57 (0.35)	2.16 (1.89)	1.03 (0.09)	0.6 (0.36)
NH-7	1.94 (1.59)	1 (0.08)	0.74 (0.35)	3.98 (2.48)	1.05 (0.06)	0.75 (0.37)
NH-8	1.69 (1.25)	0.99 (0.09)	0.54 (0.29)	2.74 (2.23)	1.04 (0.09)	0.78 (0.48)
NH-9	1.91 (1.73)	1.07 (0.11)	0.99 (0.39)	2.81 (2.1)	1.06 (0.07)	1.05 (0.47)
NH_Group	2.17 (1.72)	1.01 (0.08)	0.70 (0.39)	3.38 (2.54)	1.03 (0.08)	0.86 (0.46)

Table 4: The means and standard deviations (between brackets) of the three acoustic cues for each participant.