This is the accepted version of the paper.

This version of the publication may differ from the final published version.

FIRST LANGUAGE

VOLUME 28 – ISSUE 1 – NUMBER 84 – FEBRUARY 2008

ISSN 0142-7237
http://fla.sagepub.com

Prosodic modulation in the babble of cochlear implanted and normally hearing infants: a perceptual study using a visual analogue scale.

Journal:	*First Language*
Manuscript ID	FLA-17-0040.R3
Manuscript Type:	Original Manuscript
Keywords:	Cochlear Implants, Infants, Babble, Prosody, Prosodic modulation, Prosodic development, Perceptual experiment
Abstract:	This study investigates prosodic modulation in the spontaneous canonical babble of congenitally deaf infants with cochlear implants (CI) and normally hearing (NH) infants. Research has shown that the acoustic cues to prominence are less modulated in CI babble. However acoustic measurements of individual cues to prominence give incomplete information about prosodic modulation. In the present study, raters are asked to judge prominence since they simultaneously take into account all prosodic cues. Disyllabic utterances produced by CI and NH infants were presented to naive adult raters who had to indicate the degree and direction of prosodic modulation between syllables on a visual analogue scale. The results show that the babble of infants with CI is rated as having less prosodic modulation. Moreover, segmentally more variegated babble is rated as having more prosodic modulation. Raters do not perceive the babble to be predominantly trochaic, which indicates that the predominant stress pattern of Dutch is not yet apparent in the children's productions.

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

experiment.mov

SCHOLARONE™
Manuscripts

Abstract

This study investigates prosodic modulation in the spontaneous canonical babble of congenitally deaf infants with cochlear implants (CI) and normally hearing (NH) infants. Research has shown that the acoustic cues to prominence are less modulated in CI babble. However acoustic measurements of individual cues to prominence give incomplete information about prosodic modulation. In the present study, raters are asked to judge prominence since they simultaneously take into account all prosodic cues. Disyllabic utterances produced by CI and NH infants were presented to naive adult raters who had to indicate the degree and direction of prosodic modulation between syllables on a visual analogue scale. The results show that the babble of infants with CI is rated as having less prosodic modulation. Moreover, segmentally more variegated babble is rated as having more prosodic modulation. Raters do not perceive the babble to be predominantly trochaic, which indicates that the predominant stress pattern of Dutch is not yet apparent in the children's productions.

Acknowledgements

Introduction

In congenitally deaf infants, Cochlear Implantation (CI henceforth) in the prelexical period has been shown to have a positive effect on their speech and language acquisition (Dettman et al., 2016; Levine, Strother-Garcia, Golinkoff, & Hirsh-Pasek, 2016; Tomblin, Barker, Spencer, Zhang, & Gantz, 2005). However, the prosodic modulation of speech remains difficult even long after implantation. In Dutch prosodic modulation is achieved by manipulating three acoustic cues, viz. fundamental frequency (F0), intensity and duration (Lieberman, 1960). In perceptual terms this means that a prosodically prominent syllable in a word or prelexical utterance has higher pitch and is louder and longer. Cochlear implants are limited in transmitting the spectral and temporal information that is crucial for adequate perception of pitch (Green, Faulkner, & Rosen, 2004; Moore, 2003; O'Halpin, 2010) and intensity (Drennan & Rubinstein, 2008; Meister, Landwehr, Pyschny, Wagner, & Walger, 2011; Moore, 2003), leading to difficulties with prominence production.

For typical development there are some acoustic studies of prominence production showing that infants already start to manipulate F0, intensity and duration in canonical babble (Davis, MacNeilage, Matyear, & Powell, 2000; De Clerck, Pettinato, Verhoeven, & Gillis, 2017). This early production ability originates from infants' perceptual sensitivity to the prosodic properties of the ambient language (Friederici, Friedrich, & Christophe, 2007; Jusczyk, Cutler, & Redanz, 1993; Ramus, 2002). CI infants' speech perception is not only degraded but they also gain access to auditory input at a later chronological age. Because of this initially delayed access to the prosodic properties of their ambient language, their ability to manipulate prosodic cues is likely to be delayed as well. However, hardly any studies have focused on prelexical prominence production by early implanted infants (i.e. infants who received their CI before the age of two). The present study therefore investigates prosodic

modulation in the prelexical utterances of congenitally deaf infants who received a CI before the age of two, compared to their typically developing peers.

Prosody production by CI users

Prosody production, and more specifically word stress production, is problematic for severe-to-profoundly deaf children. They modulate prominence inadequately, resulting in anomalous intonation due to a slower speech rate (Clement, 2004; van den Dikkenberg-Pot, Koopmans-van Beinum, & Clement, 1998) as well as excessive or monotone pitch production (Clement, 2004; Kent, Osberger, Netsell, & Hustedde, 1987). Although cochlear implantation substantially improves speech perception, school-aged CI children still have problems with particular aspects of speech modulation in production, such as prominence production. Lenden and Flipsen (2007) investigated the prosodic characteristics of conversational speech of 3- to 6-year-olds with CI (implanted before age 3) by means of the Prosody-Voice Screening Profile (Shriberg et al. 1990). 83.1% of their utterances were classified as inappropriately stressed and 96.7% of these inappropriately-stressed utterances were judged to have 'excessive, equal or misplaced' stress. Carter, Dillon, and Pisoni (2002) conducted a non-word repetition task in a group of 8- to 10-year-old English children with CI (mean age at implantation 3.3 years). They found that the accuracy of imitating the stress pattern of non-words was only 61%, but there was no normally hearing (NH henceforth) control group included in the study to compare the results with. In a study with 6- to 9-year-old Dutch children with CI (age at implantation before 2 years) a non-word repetition task was performed in which the participants had to imitate the stress pattern of disyllabic non-words (Hide, 2013). It was found that CI children had a lower percentage of correctly imitated stress patterns (86% for trochaic and 81% for iambic non-words) than an age-matched NH control group (95% for trochaic and 97% for iambic non-words). Moreover, the disyllabic imitations

were classified into trochaic or iambic utterances by three raters. For the non-words that were unanimously judged as having stress on the first or the second syllable, the CI children showed less acoustic modulation than the NH children regarding pitch rises and pitch rise duration. All these studies show that school-aged CI users have difficulties with producing prosodic prominence.

Prosody production in prelexical speech

Although the production of word stress by school-aged CI users has already been investigated, studies of prosodic prominence production in prelexical speech are rare, even for typically developing infants. Davis et al. (2000) showed that prelexical NH English infants already have the necessary abilities to produce prominent syllables, but they do not yet use these in a consistent way. Raters indicated a clearly prominent syllable in less than half of the disyllabic babble. However, in those utterances with a prominent syllable F0, intensity and duration values were modulated in a way similar to adult speech. For all three cues the ratios between the more prominent and the less prominent syllables were comparable for infants and adults. One of our previous longitudinal studies followed 9 NH infants acquiring Dutch from the onset of babbling until they reached a cumulative vocabulary of 200 words (De Clerck et al., 2017). The three cues to prosody were measured in disyllabic babble and early words. It was found that from babbling onwards infants differentiate acoustically between prominent and less prominent syllables. Intensity and F0 already tend to be slightly higher in the first syllables of infant babble, but the predominant trochaic pattern was well established in early word use. Words had more prosodic modulation than babble in terms of F0 and intensity.

Research on early prominence production in CI infants is even rarer than in NH infants, therefore an acoustic study was carried out in which the three cues to prosody in disyllabic

babble and words of infants with CI and a NH control group were measured (Pettinato, De Clerck, Verhoeven, & Gillis, 2017). The results indicated that CI infants show poorer use of acoustic cues, especially of F0 and intensity. This discrepancy between CI and NH infants started from babbling onwards and the gap between the two groups became even wider when looking at early word productions. In acoustical terms, infants with CI showed less prosodic modulation from babbling onwards and the emergence of the first words did not boost prosodic modulation as was the case for NH infants.

Perceptual judgments to assess prominence production

The study of Pettinato et al. (2017) focused on the acoustic phonetics of individual prosodic cues in babble and words. Measuring F0, intensity and duration reveals subtle prosodic differences between CI and NH infants in the realisation of the isolated prosodic cues to prominence. However, a limitation of this method is that acoustically measured differences do not necessarily have a perceptual effect, meaning that raters are not necessarily able to detect these differences in prominence production. A body of research on covert contrasts in phonological development shows that significant differences between two categories of speech sounds can be measured acoustically but these differences are not perceived as such by raters (Li, Edwards, & Beckman, 2009; Scobbie, Gibbon, Hardcastle, & Fletcher, 1996). Li et al. (2009) for instance investigated the contrasts between voiceless sibilant fricatives in English and Japanese 2-to-3-year olds. Both languages have a contrast between an alveolar fricative /s/ and a postalveolar fricative (English: /ʃ/ , Japanese: /ɕ/). For four out of 22 English and two out of 21 Japanese children they found significant acoustic difference between the two types of fricatives, although a trained native transcriber transcribed all productions as /s/ (for English) or as /ɕ/ (for Japanese). Therefore, a perceptual experiment is needed to complement the acoustic findings of Pettinato et al. (2017).

An additional motivation for a perceptual study is that the prosodic cues can be in trade-off relations in speech (Lieberman, 1960), e.g. in an utterance with a prominent first syllable the measurement of one cue (e.g. F0) could for instance be higher on the first syllable, whilst the others are not (e.g. intensity and duration). Consequently the separate acoustic measurements give an ambiguous image of the realised stress pattern. Raters have been shown to make stress judgements by taking into account these trade-off relations between the different cues instead of decoding the signal into the three separate cues to prominence (Flege & Bohn, 1989; Fry, 1958). Perceptual judgements thus complement acoustic measurements because raters indicate which prosodic prominence pattern they hear by taking into account all cues at the same time.

The aim of the current study is to investigate prosodic modulation in the babble of CI and NH infants in a perceptual experiment in which naive adult raters indicate the degree and direction of prosodic differentiation in disyllabic babble on a Visual Analogue Scale (VAS). A VAS is a psychometrical measurement tool consisting of a line with two opposite speech characteristics of a stimulus at the extremes of the scale (Munson, Schellinger, & Carlson, 2012). In our study those opposites are "a very prominent first syllable" versus "a very prominent second syllable".

Canonical babble in CI infants

The focus of this study is on prosodic modulation in canonical babble of CI and NH infants. Canonical babbling is considered to be an important milestone in the construction of adult-like syllables. Canonical babble consists of consonant-vowel sequences that can be reduplicated or variegated (Oller & Eilers, 1988; Stoel-Gammon & Otomo, 1986). In

reduplicated babble the sequences follow each other repetitively (e.g. 'didi', 'mamama'). In variegated babble infants produce different segments in the syllable sequences (e.g. 'badigogu', 'kamobido'). Koopmans-van Beinum, Clement, and van den Dikkenberg-Pot (2001) propose a sensorimotor description of the prelexical speech of NH and deaf infants. They consider speech as a combination of phonation and articulation. In very early speech infants either phonate or articulate. Later on they combine phonation and articulation, resulting in the production of sequences of syllables, i.e. canonical babble. In their study Koopmans-van Beinum et al. (2001) show that deaf infants have difficulties combining phonation and articulation and thus to produce canonical babble. They argue that auditory perception and (auto-)feedback are requirements to coordinate articulatory movements and phonation of the airstream, which is essential for the production of canonical syllables.

Cochlear implantation at an early age improves speech perception and production of severely deaf infants, but studies on prelexical speech show that infants with CI continue to have difficulties with modulating the speech stream. When it comes to canonical babble, CI infants do produce more repetitive vocalisations after implantation and their reduplicated babble resembles that of younger infants with normal hearing in that the utterances consist of similar numbers of CV-repetitions (Fagan, 2015). However, Schauwers, Gillis, and Govaerts (2008) show that the intersyllabic structure of prelexical utterances remains less variegated in infants with CI: they produce more reduplicated and less variegated babble than their NH peers. NH and CI infants prefer vowel variegation over consonantal variegation and this preference is significantly stronger in the CI group. Moreover, consonant variegation in CI utterances was found to be predominantly simple (manner OR place change) rather than complex (manner AND place change) which characterised NH children's' consonant variegation. In sum, the canonical babble of CI infants is less varied or modulated at the segmental level than the

babble of NH infants. Infants with CI are able to produce reduplicated babble, but they do not

modulate the speech stream to the same extent as their NH peers in order to produce a variety

of segments. This results in less variegated babbling.

In sum, it is observed that infants with CI have less modulated prosody in acoustical terms

(Pettinato et al., 2017), and they have less modulation at the segmental level (Schauwers et al.

2008). Consequently the question arises whether CI speech is characterised by a more general

modulation issue that manifests itself both on the segmental (cf. Schauwers et al. 2008) and

the suprasegmental level. This has not yet been investigated, but present study explores the

possible influence of two maturation effects on prosody production. First of all it might be the

case that the expansion of phonological capacities influences prosodic modulation, i.e. the

phonological complexity of the babble might influence modulation. More concretely,

reduplicated babble might be prosodically less modulated than segmentally more modulated

variegated babble. Beside the expansion of phonological capacities another maturation effect

might impact prominence production. It might be the case that infants' experience with

producing canonical babble (i.e. their chronological age, counted from the onset of babbling)

leads to more prosodic modulation and/or more frequent production of the predominant

trochaic pattern.

Research questions

Two research questions are the focus of the present study:

> (1) *Is the spontaneous canonical babble of congenitally deaf infants with CI perceived to*
> *be prosodically less modulated than the babble of their NH peers?*

Given the initially delayed access to auditory input and the deprived perception of the

prosodic cues to prominence it is hypothesised that raters will perceive less modulated

prosody production in the babble of infants with CI. Moreover, the acoustic study of Pettinato et al. (2017) has shown less modulation in pitch and intensity in CI babble. It is tested whether maturation effects (i.e. phonological abilities and/or babble experience) influence prosodic modulation

(2) *Do listeners perceive the babble to be predominantly trochaic? Is there a difference*
between CI and NH infants in this respect?

The predominant stress pattern in Dutch disyllables is trochaic (Daelemans, Gillis, & Durieux, 1994). It is hypothesised that listeners will perceive more trochaic babble if infants produce the dominant prosodic characteristics of their ambient language from babbling onwards. It is tested whether maturation effects (i.e. phonological abilities and/or babble experience) influence prosodic modulation.

Method

Participants

This study was carried out in 9 infants with a CI and 9 NH infants from the CLiPS Child Language Corpus (CCLC), a collection of longitudinal audio-video data and transcriptions of 10 children with a CI and 40 NH Dutch children (Molemans, 2011; Van den Berg, 2012; Van Severen, 2012). One infant with CI was not selected for the present study because of too few recordings in the relevant age range. All children had been raised in monolingual homes acquiring Belgian Dutch.

The children with CI were recruited from an ENT-unit of the St-Augustinus hospital in Antwerp/Belgium in 2000-2001 (Schauwers, 2006). The participants had all been diagnosed with a profound congenital hearing loss through a neonatal hearing screening during the first weeks of life. No other health or developmental problems were apparent. The children received a multichannel Nucleus-24 CI (Cochlear Corp., Sydney, Australia). This device consists of 22 intra-cochlear electrodes, like more recent CIs. The data and results in the present study are still representative since the present-day electrodes are not significantly improved in transmitting spectro-temporal information.

All infants were implanted before 20 months, with age of implantation ranging from 5 to 19 months (M = 12 months; SD = 5 months). Before implantation the average hearing threshold, i.e. Pure Tone Average (PTA), was 113 dBHL and the range was 93-120 dBHL (SD = 9 dBHL). One year after implantation the PTA steadily decreased to 30-52 dBHL (M = 40 dBHL; SD = 7 dBHL). All recordings used in this study were made while the children were unilaterally implanted. The auditory characteristics are listed in Table 1.

INSERT <u>TABLE 1</u> ABOUT HERE

As a control group for this study, 9 NH children were randomly selected from the corpus. The

infants had been recruited from day-care centres, families known by the researchers and by advertisements. 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 Flemish infant welfare centre (Molemans, 2011). Normal language development had been verified by means of the Dutch version of the CDI ("N-CDI", Zink & Lejaegere, 2001). The N-CDI was filled out by the parents of the NH children to estimate 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.

For the present experiment babble data from both groups was selected from the onset of babbling until the children reached a cumulative vocabulary level of 200 words (cf. 'experimental stimuli'). This cut-off point was arbitrary but motivated by the fact that infants do not stop babbling after the onset of word use and because the selected time range provided enough data for every child included in this study. This procedure implies that the chronological age of each child at the beginning of data selection and at the end of data selection was (potentially) different. The onset of babbling was determined by a True Canonical Babbling Ratio (tCBR) of 0.15 (Molemans, Van den Berg, Van Severen, & Gillis, 2011; Oller & Eilers, 1988). The tCBR is the ratio of the syllables with 'true consonants' (i.e. all consonants except glottals (/h/, glotal stop) and glides (/w/, /j/)) over all syllables produced. The mean age of the CI children at the start of the recordings used in the present study was 18 months (*SD* = 5 months). The mean age at the cut-off point was 27 months (*SD* = 5 months). The mean age of the NH children at the onset of recordings was 8 months (*SD* = 2 months). The mean age at the cut-off point was 22 months (*SD* = 2 months). The ages of the individual children at the time of recording is given in Table 2.

INSERT <u>TABLE 2</u> ABOUT HERE

Perceptual experiment

A perceptual experiment was set up in which naive adult raters indicated the degree and direction of prosodic differentiation in disyllabic babble on a VAS.

Adult raters. 30 naive native Dutch-speaking adults (mean age 23 years, range 18-42 years) made the prominence judgements in the perceptual experiment. The raters were naive since they were not informed about the purpose of the study. Moreover, they were not familiar with CI and NH children's speech. None of the raters reported problems with hearing or health, nor developmental disorders.

Experimental stimuli. For the present study, all eligible disyllabic utterances were selected from the CCLC. Each recording lasted 60-90 minutes and the portions of the recordings in which the child was vocally most active were selected for further processing. These selections were restricted to approximately 20 minutes. From the selections only disyllabic babbled utterances were selected when they consisted of two vocalic phases minimally separated by a clear consonantal phase. Moreover, disyllabic babble was included if there was no concurrent speech or noise and if they were not produced with a creaky, breathy, excessively loud or whispery voice. A total of 524 disyllabic babble (165 CI babble, 359 NH babble) was included as stimuli for the perceptual experiment (see Table 2).

Since maturation effects might influence the production prosodic modulation, two variables were investigated in this study. The first maturation variable is the infants' babble experience. Since infants start babbling at different ages and since there is a large discrepancy between

the chronological age at onset of babbling of the CI and NH infants, it is more meaningful to take into account the chronological age counted from the onset of babbling onwards. More specifically, the month in which the child started babbling was labelled as 0, the first month after the onset of babbling as 1, the second month as 2, etc. The second maturation variable is the phonological complexity of the babble. The canonical babble utterances were categorised into reduplicated and variegated babble. The prelexical utterances in the recordings of the CCLC are transcribed in broad phonemic categories (detailed information on the transcription is provided in Molemans, 2011; Van den Berg, 2012; Van Severen, 2012). These broad phonemic transcriptions were used to automatically categorise babble into variegated or reduplicated babble. When the same segmental content was repeated in the two syllables the utterance was classified as reduplicated (e.g. '*baba*'). Whereas in variegated babble the syllables differ at the segmental level by consonantal and/or vocalic variation (e.g. '*baka*'). This means that there were two intermediary steps: the sound files were first transcribed as part of a previous research project (Molemans, 2011; Van den Berg, 2012; Van Severen, 2012), and then these transcriptions were automatically coded as reduplicated or variegated babble. 44% of the disyllabic babble included in the present study was reduplicated babble (reduplicated CI babble: 45%, reduplicated NH babble 43%) and 56% of the stimuli was variegated babble (variegated CI babble: 55%, variegated NH babble 57%). Table 2 contains an overview of the number of reduplicated and variegated babble per child.

As a reliability check 20% of the automatically categorised stimuli (106 babble: 53 NH and 53 CI utterances) was manually re-categorised into reduplicated and variegated babble by the first author. The raw sound files were used to judge whether the babble has repeated segmental content or not. In 91 out of 106 utterances categorisation was identical, resulting in an agreement of 86% and a Cohen's Kappa of 0.68.

Experimental design. In the rating task, the raters indicated the more prominent syllable. Moreover, by moving the slider on the VAS they indicated the relative degree of prominence of the syllable. Studies have shown that perceptual ratings on a VAS are informative for capturing the perceptual effects of fine phonetic detail in child speech (Julien & Munson, 2012; McAllister Byun, Harel, Halpin, & Szeredi, 2016; Munson, Schellinger et al., 2012). Munson and Carlson (2016) showed that a VAS is a useful method for experimental and observational studies of within-category speech sound perception as it allows gradual responses.

The VAS used in the present study was a sliding bar of which a screencast can be found in supplemental material 1 and a print screen of three different rating positions can be found in Figure 1. Two circles above the sliding bar served as a visualisation of the two syllables of each stimulus. The initial position of the slider was the midpoint of the VAS (Figure 1a). When sliding to the left, the left circle became larger while the right one became smaller, indicating a more prominent first syllable (Figure 1b). In order to indicate a more prominent second syllable, the slider was moved to the right, creating a larger right circle and a smaller left circle (Figure 1c). The position of the slider moved along a scale ranging from 0 (extreme left) to 100 (extreme right). Thus, a range between 0 to 49 means a more prominent first syllable (the lower the number, the more prominent the first syllable). Point 50 indicates equally prominent syllables, and a position in the range from 51 to 100 indicates a more prominent second syllable (the higher the number, the more prominent the second syllable).

INSERT FIGURE 1 ABOUT HERE

Experimental procedure. Before the start of the experiment, the raters went through an information phase and a familiarisation phase. During the information phase, the purpose of the experiment was explained in a section with instructions (see supplemental material 2) and raters were told that they could replay an utterance 2 times (i.e. each utterance could be played maximally 3 times).

The familiarisation phase consisted of an example and trial phase. During the example phase utterances with clear prominence relationships (i.e. clear equal prominence or a clearly prominent first or second syllable) were presented. The raters saw an animation of 6 clear examples of how the sliding bar was moved to the appropriate position. During the trial phase the raters had to indicate the prominence on similarly clear utterances. The task in the trial phase was the same as in the actual experiment, i.e. "indicate the prosodic modulation and stress pattern on the VAS". The description and design of the example and trial utterances can be found in supplemental material 3. In the actual experiment 524 disyllabic babble utterances produced by CI and NH infants were presented in random order to the raters. The total duration of the experiment was approximately 60 minutes. For every stimulus we recorded the rate and how many times raters listened to the stimulus.

The experimental interface was created by means of Functional Reactive Programming (Czaplicki, 2012), a tool developed for creating responsive graphical interfaces. The programming language used to create the front end was Elm, and Haskell was used for the back end. The experiment was designed as an online application and ran on an iMac (2,9 GHz Intel Core i5, on a 21.5 inch screen) in a quiet room for all 30 participants. During the experiment the selected disyllabic babble utterances were presented one by one to the raters via SONY MDR-1R headphones. Every stimulus was preceded and followed by one second

of surrounding silence.

Statistical approach

Generalised mixed models. To analyse the data generalised mixed models (GLMM) (Baayen, 2008) were run in R (R Core Team, 2013) with the lme4 package (Bates, Maechler, Bolker, & Walker, 2014). GLMM is an appropriate tool to examine data that are hierarchically structured: the stimuli in the present study ($n = 524$) are embedded in different infants ($n = 18$) and all utterances are rated by different raters ($n = 30$). Moreover, GLMM is robust to missing data and different numbers of data points per participant. These models consist of two parts: a random and a fixed part. The random part takes into account the variation caused by the random effects, e.g. the variation between individuals. The fixed part consists of the independent variables that may have an effect on the dependent variables, i.e. participant group (CI or NH babble), utterance type (variegated or reduplicated babble) and age at onset of babbling.

The statistical models are constructed iteratively: first the random effects are added one by one and then the fixed effects are added one at a time. After adding a random or fixed effect the new model is compared to the previous model by means of likelihood ratio tests. Only the effects that improve the fit of the model are included. The best fitting model explains the largest amount of variance with the smallest set of predictors.

The dependent variables. The ratings on the VAS resulted in a continuous score from 0-100, representing the degree of prosodic modulation. A set of analyses with this continuous

dependent variable were conducted (see supplemental material 4). However, in neither of the analyses were the residuals normally distributed, meaning that a GLMM analysis of the continuous variable was not permitted. As an alternative the data were recoded into binomial dummy variables that are informative of the rating zone on the VAS instead of using the raw score (from 0 to 100) as dependent variable.

As shown in Figure 2, the collected ratings form a multimodal distribution: a distribution on the left and right side of the VAS and a peak at the midpoint of the VAS. These variables were 'trochee', 'iamb', 'midpoint'. 'Trochee' (score: 0-49) represents the utterances that are rated as having a prominent first syllable, which is the predominant Dutch pattern. 'Iamb' (score: 51-100) represents the utterances that are rated as having a prominent second syllable. 'Level' (score: 50) represents the utterances that are rated at the midpoint of the axis. The midpoint was also the default position of the sliding bar before being moved by the raters. When an utterance is rated at the midpoint this is an indication that there is no clear prosodic differentiation between syllables.

<u>INSERT FIGURE 2 ABOUT HERE</u>

The binomial dependent variables were analysed by means of Logistic Regression Analyses in the form of GLMM. The estimates (E), standard errors (S.E.), t- and p-values of the fixed effects of the four best fitting models (i.e. one for every dependent variable) are reported in the results section. These values are computed in logits, but to facilitate the interpretation they are converted to the likelihood or probability (expressed in percentages) to be rated in this area of the VAS.

 The aim of the analyses was to investigate for every rating zone whether (1) CI utterances were more likely to be rated in the respective zone as compared to NH utterances and vice versa, (2) chronological age at onset of babbling has an impact and (3) whether reduplicated

babble is more likely to be rated in a particular zone as compared to variegated babble and vice versa.

The first research question focuses on the prosodic modulation in the babble of CI and NH infants. Three analyses were performed to answer this research question. The first analysis looked at the probability of an utterance being rated at the midpoint of the VAS (score: 50). Since the midpoint of the VAS is considered to be the zone that represents the least prosodic modulation, this analysis aims to answer the question whether the babble of infants with CI is more likely to be rated in this zone. The analysis controls for a possible effect of babble experience (i.e. chronological age counted from the onset of babbling) and phonological complexity (i.e. the babble type: reduplicated versus variegated babble) on prosodic modulation.

The second analysis that investigates raters' perception of prosodic modulation focussed on those utterances that were indicated as having the most modulated prominence. This analysis aims to answer the question whether the utterances of the NH group are more likely to be rated as having a clearly prominent syllable than the utterances of infants with CI. Again age at onset of babbling and the babble type are added as predicting variables.

The utterances that are rated more towards the extremes of the VAS are considered to have the most prosodic modulation. However, looking at the extremes of the raw VAS (e.g. the stimuli rated in zone 0-20 and zone 80-100) would not be informative since individual raters may well apply different strategies to indicate prominence on a VAS: some use the entire scale (from 0 to 100), but most of them do not use the entire VAS and tend to centre their ratings in a smaller range (e.g. from 19 to 82). Therefore it is more informative convert the

ratings of every rater to z-scores. The z-scores are used to determine the extreme ratings per individual rater. A stepwise approach was used to determine these extreme ratings. First of all the range of the ratings of each rater was determined: e.g. for rater 1 the leftmost z-score was -2.33, whereas the rightmost rating was a z-score of 3.01. In a next step this range was divided into 5 equal parts in order to determine the 20% leftmost and 20% rightmost parts of the VAS for each individual rater: e.g. for rater 1 the extremes range from -2.33 and -1.26 (left) and 1.94 and 3.01 (right). The utterances rated in the first and fifth part of the selected range were coded as being rated at the extremes of the scale: e.g. for rater 1 the left extreme included 45 utterances and the right extreme 24 utterances. The dependent variable in the third analysis was the number of times an utterance was replayed by the raters. If the prosodic modulation of a stimulus is hard to judge, raters are expected to replay it more often than when prominence is clear.

The second research question was whether listeners perceive the babble to be predominantly trochaic and whether there is a difference between CI and NH infants in this respect. This was investigated by two analyses. The first analysis looked at the probability that an utterance was rated on the left, trochaic side (score: 0-49) of the VAS. The second analyses looked at the probability that an utterance was rated on the right, iambic side (score: 51-100) of the VAS. Both analyses control for possible maturation effects by adding age at onset of babbling and utterance type to the models.

Results

Analyses of raters' perception of prosodic modulation in babble of CI and NH infants

The first set of analyses aims to answer the first research question: Are spontaneous canonical babble of congenitally deaf infants with CI perceived to be prosodically less modulated than the babble of their NH peers?

Midpoint. The first analysis looked at the probability of an utterance being rated at the midpoint of the VAS (score: 50). Since the midpoint of the VAS is considered to be the zone that represents the least prosodic modulation, this analysis aims to answer the question whether the babble of infants with CI is more likely to be rated in this zone. Moreover, it is investigated whether maturation variables (i.e. utterance type and babble experience) influence the likelihood of babble being rated with no modulation.

The best model with "midpoint" as dependent variable had participant group and utterance type as fixed effects (supplemental material 5). Adding age at onset of babbling did not improve the fit model. The random part controlled for the variance of the different infants, utterances and raters. The intercept of the model was estimated at -3.222 logits or 4% (*S.E.* = 0.297, *z* = -10.833, *p* < 0.001), meaning that babble is significantly less likely to be rated at the midpoint, than anywhere else on the VAS. This finding is not surprising since this variable focuses on only one point on the VAS, in comparison to all other points. This finding indicates that the cursor was very likely being moved when making a judgement on the prominence pattern of babble. Regarding the question whether there is a difference between CI and NH infants, the results show that CI utterances are significantly more likely to be rated at the midpoint: reduplicated CI babble utterance has a probability of 3.5% of being rated at

the midpoint, whereas the probability is only 2.6% for reduplicated NH utterance (-3.222 - 0.401 logits; E = -0.401, $S.E.$ = 0.188, z = -2.127, p = 0.033). Reduplicated babble of CI infants is judged as having less modulated prominence. The same holds for variegated babble. Variegated CI babble utterance has a probability of 2.2% (-3.222 - 0.527 logits; E = -0.527, $S.E.$ = 0.136, z = -3.878, p < 0.001) to be rated at the midpoint and that of a NH infant has a probability of 1.5% (-3.222 - 0.401 + 0.527 logits) to be rated at the midpoint. All these differences are shown to be significant. These results indicate that babble of CI infants is significantly more likely judged as having no clearly prominent syllable. A similar result is found for the difference between variegated and reduplicated babble. In both infant groups reduplicated babble had a higher probability (i.e. 3.5% for a CI- and 2.6% for a NH-reduplicated babble) to be rated at the midpoint compared to segmentally variegated babble (i.e. 2.2% for a CI- and 1.5% for a NH-reduplicated babble). The age at onset of babbling did not improve the fit of the statistical model, meaning that it did not impact the probability of being rated with no clearly stressed syllable.

It can be concluded that babble of CI children is more likely to be rated as having no clearly prominent syllable. Regarding the effect of the maturation variables it can be concluded that phonetic complexity influences prosodic modulation: reduplicated babble is more likely to be rated as having no clearly prominent syllable and thus as having less prosodic modulation, irrespective of the infant group or the age at onset of babbling at which they are produced.

Extremes. The best model with "extreme" as dependent variable had participant group and utterance type as fixed effects (Supplemental material 6). Again the random part controlled for the variance of the different infants, utterances and raters and the fixed effects were participant group and babble type. The intercept of the model was estimated at -0.826 logits

or 30% (*S.E.* = 0.116, z = -7.133, p < 0.001), indicating that babble is significantly less likely to be rated at the extremes than anywhere else on the VAS. Regarding the question whether there is a difference between CI and NH infants in prosodic modulation in babble, the results indeed show that NH utterances are significantly more likely to be rated at the extremes: there is a probability of 30% for reduplicated babble of a CI infant being rated at the extremes, whereas the probability that reduplicated NH babble is rated at the extremes is 36% (-0.826 + 0.255 logits; E = 0.255, *S.E.* = 0.129, z = 1.981, p = 0.048). When looking at variegated utterances it appears that variegated CI babble has a probability of 38% (-0.826 + 0.308 logits; E = 0.308, *S.E.* = 0.084, z = 3.667, p < 0.001) and that variegated NH babble has a probability of 43% (-0.826 + 0.255 + 0.308 logits) to be rated at the extremes. All these differences are shown to be significant. The age at onset of babbling has no significant effect on the probability of being rated at the extremes (E = 0.004, *S.E.* = 0.010, z = 0.426, p < 0.670).

It can be concluded that the canonical utterances of the NH infants are more likely to be rated at the extremes of the ratings compared to the CI infants' utterances. This is an indication that the prosodic modulation within NH babble is more salient. This difference between groups is apparent in both variegated and reduplicated babble. Regarding the effect of the maturation variables it can be concluded that phonetic complexity influences prosodic modulation: variegated babble is more likely to be rated at the extremes than reduplicated utterances irrespective of the infant group or the age at onset of babbling at which they are produced.

Analysis of number of replays. A third analysis investigates how many times an utterance was replayed by the raters. The random effects again account for the variance caused by the different infants, raters and stimuli. The fixed effects in this analysis were infant group,

babble type and extreme (i.e. part of the extreme ratings or not). The fixed effects in the best

fitting model were participant group and the variable 'extreme' (i.e. the utterance is rated at

one of the extremes or not). Adding babble type did not improve the fit of the model, meaning

that there was no significant difference in the number of replays of reduplicated versus

variegated babble. The results of this analysis are displayed in supplemental material 7.

The intercept of this statistical model was estimated at 0.423 (*S.E.* = 0.046, *t* = 9.156, *p* <

0.001), meaning that an average CI utterance that is rated anywhere at the VAS except at the

extremes is replayed on average 0.423 times. Utterances of the NH infants are repeated

significantly less than those of the infants with CI (*E* = -0.058, *S.E.* = 0.017, *t* = -3.349, *p* =

0.004), indicating that it was easier to judge the prominence of NH babble than CI babble.

The stimuli rated at the extremes were repeated significantly less than those at the rest of the

VAS (*E* = -0.126, *S.E.* = 0.009, *t* = -14.613, *p* < 0.001). The age at onset of babbling did not

improve the fit of the statistical model, meaning that it did not impact the number of replays.

These results are supporting evidence for the findings of the VAS analysis that showed that

NH stimuli are rated as having clearer prominence and as those utterances that were rated at

the ends of the continuum are repeated less often and are thus easier to judge.

Analyses of raters' perception of the trochaic pattern in babble of CI and NH infants

The following two analyses aim to answer the second research question: Do listeners perceive

the predominant trochaic stress pattern in babble? Is there a difference between CI and NH

infants in this respect?

Trochee. The first analysis looked at the probability that an utterance was rated on the left,

trochaic side (score: 0-49) of the VAS. The random part of the best fitting model controlled

for the variance explained by participant identity (i.e. the 18 infants), utterance identity (i.e. a

unique number for each of the 524 babble utterances) and rater identity (i.e. the 30 adult

raters). The only fixed effect required in the best fitting model was infant group (CI versus

NH). Adding babble type (reduplicated versus variegated babble) and age at onset of babbling

as fixed effects did not improve the fit of the model. An overview of the best fitting model

with 'trochee' as dependent variable can be found in supplemental material 8.

The intercept of the model was estimated at 0.157 logits or a likelihood of 54% (*S.E.* =

0.268, *z* = 0.586, *p* = 0.558*)*. This positive intercept indicates a higher probability for

utterances to be rated as a trochee than at any other zone on the VAS. However this effect is

not significant. There is no significant difference between the ratings of the CI and NH babble

(*E* = -0.041, *S.E.* = 0.345, *z* = -0.120, *p* = 0.904). These results show that babble is not

significantly more likely to be rated at the trochaic side of VAS, nor that the NH infants are

more likely to produce trochees than CI infants or vice versa. The babble type did not

improve the fit of the statistical model, indicating that the segmental variation of the utterance

does not lead to more ratings on the trochaic side of the VAS. Age at onset of babbling did

not improve the fit of the model either, meaning that there is no age effect on the probability

to be rated as a trochee. Regarding the second research question it can be concluded that the

predominant trochaic pattern is not significantly more often perceived by raters and this is the

case for both infant groups.

Iamb. The aim of the second analysis was to see whether there is a difference between groups

and babble types regarding the probability to be rated on the right side of the VAS (score: 51-

100). The same fixed and random effects as in the analysis with trochee as dependent variable

were added. The statistical model with the best fit (supplemental material 9) had participant

identity, utterance identity and rater identity as random effects and participant group and

utterance type as fixed effects. The intercept was estimated at -0.929 or a likelihood of 28% (*S.E.* = 0.320, *z* = -2.905, *p* = 0.004). This significant negative intercept indicates that utterances are less likely to be rated at the right side of the VAS than at all the other rating zones together. As for the analysis of trochees there was no significant difference between the ratings of the CI and NH babble (*E* = 0.094, *S.E.* = 0.389, *z* = 0.243, *p* = 0.808) suggesting that there is no difference between groups in the probability that an utterance is rated as having a more prominent second syllable. The analysis showed that variegated babble is significantly more likely to occur at the iambic side of the VAS (*E* = 0.401, *S.E.* = 0.192, *z* = 2.088, *p* = 0.037). For the CI infants the likelihood of reduplicated babble occurring at the iambic side of the VAS is 28% (-0.929 logits) whereas it is 37% for variegated babble (-0.929 + 0.401 logits). For the NH infants the likelihood of reduplicated babble occurring at the iambic side of the VAS is 30% (-0.929+0.094+0.401 logits) and 39% for variegated babble (-0.929 + 0.401 logits). It can be concluded that variegated babble is more likely to be rated as having a more prominent second syllable than babble with a repetitive segmental content, irrespective of the children's hearing status.

Discussion

This paper examines the prosodic modulation in the babble of congenitally deaf infants with CI and their NH peers. Two previous studies of the infants included in the present study suggested that early implanted infants with CI may have difficulties with modulating speech from babbling onwards: one study showed that CI infants' babble is segmentally less modulated (Schauwers et al., 2008) and the other study showed that CI babble is prosodically less modulated (Pettinato et al., 2017). Schauwers et al. (2008) showed that infants with CI produce less variegated and more reduplicated utterances than their NH peers. Moreover, there is less segmental variation in variegated babble of CI infants. Regarding the modulation of the acoustic cues to prosodic prominence, Pettinato et al. (2017) showed that infants with CI made smaller differentiation between the two syllables of their babble utterances in terms of pitch and intensity, whereas there was no significant difference in duration differentiation in babble. However, since the acoustic cues were measured separately in this study and since trade-off relations between the three prosodic cues are expected to occur (Lieberman, 1960), the current study aimed to take into account the relations between separate cues by presenting babble to raters in a perceptual rating task. This was done by means of a VAS on which naive adults had to indicate the degree and direction of differentiation between the two syllables of babble. The click location on the VAS was analysed in order to find an answer to the following research questions:

(1) Are spontaneous canonical babble of congenitally deaf infants with CI perceived to be prosodically less modulated than the babble of their NH peers?

(2) Do listeners perceive babble to be predominantly trochaic? Is there a difference between CI and NH infants in this respect?

Regarding the first research question, results of this study showed that babble of the CI infants

was indeed perceived as less prosodically modulated. This was shown in the analyses of the rating position on the VAS and confirmed by the analysis of the number of replays. First of all, the analyses of the VAS ratings revealed that the babble of CI infants was more likely to be scored at the midpoint of the VAS: this is the default position of the slider which represents equally prominent syllables. Therefore, this result indicates that raters found it more difficult to decide which syllable was more prominent in CI babble than in NH babble. Secondly, the VAS ratings showed that CI babble was less likely to be scored towards the extremes of the rating scale. Utterances that are rated at the extremes are considered to have the most prosodic modulation: one of the two syllables is clearly prominent. The higher likelihood of CI babble being rated towards the centre of the VAS rather than towards the extremes of the scale is consistent with a perception of smaller prosodic modulation in the canonical babble of CI infants. Interestingly, the analysis of the number of replays as dependent variable has also shown that raters listened significantly more often to CI babble than to NH babble. This is also suggestive of the greater difficulty in judging which syllable is more prominent in CI utterances.

The result that adult raters perceived CI babble as being less modulated is consistent with the literature, which shows anomalous prominence production in school-aged children with cochlear implants (Carter et al., 2002; Hide, 2013; Lenden & Flipsen, 2007). Moreover, the results from the present perceptual experiment support the previous acoustic findings that pitch and intensity are less modulated in CI babble (Pettinato et al., 2017). The present study shows that gradient ratings on a VAS reveal subtle differences in prosodic modulation in infant babble and therefore reveal modulation differences between a typical group and a clinical group (CI group in this case). This finding confirms the studies that show the importance and validity of perceptual judgements (on a VAS) to track subtle phonetic

differences (McAllister Byun et al., 2016; Munson & Payesteh, 2013; Munson et al., 2012).

The analyses also controlled for the influence of possible maturation effects on prosodic modulation in canonical babbling. The analyses showed that an infant's experience with babbling did not boost prosodic modulation, i.e. babble produced shortly after the onset of babbling was not less modulated than babble produced later and vice versa. On the other hand, the infants' phonological capacities did influence prosodic modulation. The analysis showed that phonologically more complex, variegated, babble is significantly more likely to be rated towards the extremes of the VAS than towards the midpoint of the scale: this means that variegated babble has more prosodic modulation. No interaction effect between the infant groups and utterance types was found, which indicates that variegated babble is rated as being more modulated in both groups. Since infants with CI produce less variegated babble in comparison to NH infants (Schauwers et al., 2008) it is possible that prosodic modulation in the babble of CI infants is linked to reduced segmental variegation in their utterances. It should be noted that this study is the first to explore this link between segmental and suprasegmental modulation. Therefore the present result is interpreted as an exploratory finding that needs replication in a study that is designed to investigate this link.

A plausible explanation for the indirect link between the lack of modulation at the segmental and at the suprasegmental level is that (prelexical) infants with CI have difficulties modulating the speech stream, due to their reduced perception of the speech signal. Infants with CI perceive less variation in the speech signal due to the limited transmission of spectro-temporal information. The results of the present study suggest that this reduced perception of variation is also reflected in reduced production of variation, both at the segmental and the

suprasegmental level. Moreover, the degraded perception of the speech stream leads to storage deficits in working memory (Nittrouer, Caldwell-Tarr, & Lowenstein, 2013). Children with more limited working memory in turn are less efficient in processing stress (Torppa et al., 2014). The results of the present study suggest that perception and processing problems in CI infants not only impact word stress discrimination but also word stress production. More specifically, a reasonable hypothesis is that the workload involved in producing two segmentally and/or prosodically equal syllables (i.e. reduplicated babble and equally prominent syllables) is lower than the workload required for producing modulated syllables (i.e. variegated babble and babble utterances with a prominent syllable). This hypothesis suggested by the results of the present study requires further experimental testing in order to figure out whether children with poorer working memory indeed have more modulation issues at the segmental as well as the suprasegmental level.

The second research question was whether listeners perceive the predominant trochaic pattern in the babble of infants and whether there is a difference between CI and NH infants in this respect. Studies on early prominence production often focus on the occurrence of characteristics of the ambient stress pattern (Davis et al., 2000; Hallé, De Boysson-Bardies, & Vihman, 1991). In the present paper raters indicated whether the utterances were trochaic (i.e. the ambient Dutch pattern), iambic or equally stressed. The results indicate that naive adult raters did not show a trochaic or an iambic bias in their judgements. Moreover, the results did not show a difference between CI and NH babble in this analysis. Previous acoustic studies (De Clerck et al., 2017; Pettinato et al., 2017) showed that intensity and F0 tend to be higher in the first syllables of infant babble, thus suggesting a trochaic pattern. The predominant trochaic pattern became more dominant than the iambic pattern in early word use. However this tendency is not found in the present perceptual study. It might well be the case that raters

find it substantially more difficult to judge the stress pattern of meaningless babble compared to meaningful first words. Although several studies have found ambient language effects in the babbling stage (Boysson-Bardies, Hallé, Sagart & Durant, 1989; Boysson-Bardies & Vihman, 1991; Whalen, Levitt and Wang, 1991), Engstrand, Williams, and Lacerda (2003) for instance have shown that the detection of ambient-language effects in the babble utterances of 18 month-olds becomes substantially easier for trained raters when the dataset includes possible words and imitations. Similarly, the previous study on NH infants has shown enhanced prosodic differentiation in words as compared to babble (De Clerck et al., 2017). The ambient trochaic pattern thus may only be detectable from word-use onwards. Therefore it would be interesting to compare the present results to the results of a rating task in which early lexical utterances are judged.

In conclusion, the present study shows that perceptual judgements by naïve raters reveal subtle prosodic differences between the canonical babble of congenitally deaf infants with CI and NH infants. The raters indicated less prosodic modulation in CI babble. Interestingly, segmentally more variegated utterances are rated as having more prosodic modulation. Listeners did not perceive a bias towards the predominant trochaic pattern.

References

Baayen, H. (2008). *Analyzing linguistic data: a practical introduction to statistics*. Cambridge, UK: Cambridge University Press.

Bates, D., Maechler, M., Bolker, B.M. , & Walker, S. (2014). lme4: linear mixed-effects models using Eigen and S4. R package version 1.1-7. Retrieved from http://CRAN.R-project.org/package=lme4

De Boysson-Bardies, B., Halle, P., Sagart, L., & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language, 6*, 1-17.

De Boysson-Bardies, B., & Vihman, M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language, 67*, 297-319.

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

Clement, C. J. (2004). *Development of vocalizations in deaf and normally hearing children.* (Doctoral dissertation), Universiteit van Amsterdam, Utrecht, The Netherlands.

Czaplicki, E. (2012). *Elm: concurrent FRP for functional GUIs.* Harvard University, Cambridge, UK.

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

Davis, B., MacNeilage, P., Matyear, C., & Powell, J. (2000). Prosodic correlates of stress in babbling: an acoustical study. *Child Development, 71*, 1258-1270.

De Clerck, I., Pettinato, M., Verhoeven, J. , & Gillis, S. (2017). Is prosodic production driven by lexical development? Longitudinal evidence from babble and words. *Journal of Child Language, 44*(5), 1248-1273.

Dettman, S., Dowell, R., Choo, D., Arnott, W., Abrahams, Y., Davis, A., . . . Briggs, R. (2016). Long-term communication outcomes for children receiving cochlear implants younger than 12 months: a multicenter study. *Otology & Neurotology, 37*(2), e82-e95.

Drennan, W. R., & Rubinstein, J. T. (2008). Music perception in cochlear implant users and its relationship with psychophysical capabilities. *Journal of Rehabilitation Research and Development, 45*(5), 779-789.

Engstrand, O., Williams, K., & Lacerda, F. (2003). Does babbling sound native? Listener responses to vocalizations produced by Swedish and American 12- and 18-month-olds. *Phonetica, 60*(1), 17-44.

Fagan, M. K. (2015). Why repetition? Repetitive babbling, auditory feedback, and cochlear implantation. *Journal of Experimental Child Psychology, 137*, 125-136.

Flege, J. E., & Bohn, O. S. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition, 11*, 35-62.

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

Fry, D. B. (1958). Experiments in the perception of stress. *Language & Speech, 1*, 126-152.

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

Hallé, P., De Boysson-Bardies, B., & Vihman, M. (1991). Beginnings of prosodic organization: intonation and duration patterns of disyllables produced by Japanese and French infants. *Language & Speech, 34*(4), 299-318.

Hide, O. (2013). *Acoustic features of speech by young cochlear implant users. A comparison with normal-hearing and hearing-aided age mates.* (Doctoral dissertation), University of Antwerp, Antwerp, Belgium.

Julien, H., & Munson, B. (2012). Modifying speech to children based on their perceived phonetic accuracy. *Journal of Speech, Language, & Hearing Research, 55*, 1836-1849.

Jusczyk, P., Cutler, A., & Redanz, N. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*, 675-687.

Kent, R. D., Osberger, M. J., Netsell, R., & Hustedde, C. G. (1987). Phonetic development in identical twins differing in auditory function. *Journal of Speech & Hearing Disorders, 52*(1), 64-75.

Koopmans-van Beinum, F., Clement, C., & van den Dikkenberg-Pot, I. (2001). Babbling and the lack of auditory speech perception: a matter of coordination? *Developmental Science, 4:1*, 61-70.

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

Levine, D., Strother-Garcia, K., Golinkoff, R. M., & Hirsh-Pasek, K. (2016). Language development in the first year of life: what deaf children might be missing before cochlear implantation. *Otology & Neurotology, 37*(2), e56-e62.

Li, F., Edwards, J., & Beckman, M. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics, 37,* 111-124.

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

McAllister Byun, T., Harel, D., Halpin, P. F., & Szeredi, D. (2016). Deriving gradient measures of child speech from crowdsourced ratings. *Journal of Communication Disorders, 64*, 91-102.

Meister, H., Landwehr, M., Pyschny, V., Wagner, P., & Walger, M. (2011). The perception of sentence stress in cochlear implant recipients. *Ear & Hearing, 32*(4), 459-467.

Molemans, I. (2011). *Sounds like babbling: A longitudinal investigation of aspects of the prelexical speech repertoire in young children acquiring Dutch.* (Doctoral dissertation), University of Antwerp, Antwerp, Belgium.

Molemans, I., Van den Berg, R., Van Severen, L., & Gillis, S. (2011). How to measure the onset of babbling reliably? *Journal of Child Langage*, 1-30.

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), 243-254.

Munson, B., Bjorum, E., & Windsor, J. (2003). Acoustic and perceptual correlates of stress in nonwords produces by children with suspected developmental apraxia of speech and children with phonological disorder. *Journal of Speech, Language, & Hearing Research, 46*, 189-202.

Munson, B., & Carlson, K. U. (2016). An exploration of methods for rating children's productions of sibilant fricatives. *Journal of Speech, Language & Hearing Research, 19*(1), 36-45.

Munson, B., & Payesteh, B. (2013). *Clinical feasibility of visual analog scaling.* Paper presented at the 2013 Convention of the American Speech-Language Hearing Association.

Munson, B., Schellinger, S. K., & Carlson, K. U. (2012). Measuring speech-sound learning using visual analog scaling. *SIG 1 Perspectives on Language Learning & Education, 19*, 19-30.

Nittrouer, S., Caldwell-Tarr, A., & Lowenstein, J. H. (2013). Working memory in children with cochlear implants: problems are in storage, not processing. *International Journal of Pediatric Otorhinolaryngology, 77*(11), 1886-1898.

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

Oller, D., & Eilers, R. (1988). The role of audition in infant babbling. *Child Development, 59*(22), 441-449.

Pettinato, M., De Clerck, I., Verhoeven, J., & Gillis, S. (2017). Expansion of prosodic abilities at the transition from babble to words: a comparison between children with cochlear implants and normally hearing children. *Ear & Hearing, 38*(4), 475-486.

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computin. Retrieved from http://www.R-project.org

Ramus, F. (2002). Language discrimination by newborns: teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition, 2*, 85-115.

Schauwers, K. (2006). *Early speech and language development in deaf children with a cochlear implant: A longitudinal investigation.* (Doctoral dissertation), University of Antwerp, Antwerp, Belgium.

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

Scobbie, J. M., Gibbon, F., Hardcastle, W. J., & Fletcher, P. (1996). Covert contrast as a stage in the acquisition of phonetics and phonology. *QMC Working Papers in Speech and Language Sciences, 1*, 13-62.

Stoel-Gammon, C., & Otomo, K. (1986). Babbling development of hearing impaired and normally hearing subjects. *Journal of Speech & Hearing Disorders, 51*, 33-41.

Tomblin, J., Barker, B., Spencer, L., Zhang, X., & Gantz, B. (2005). The effect of age at cochlear implant initial stimulation on expressive language growth in infants and toddlers. *Journal of Speech, Language, & Hearing Research, 48*, 853-867.

Torppa, R., Faulkner, A., Huotilainen, M., Jarvikivi, 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), 182-191.

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.* (Doctoral dissertation), University of Antwerp, Antwerp, Belgium.

van den Dikkenberg-Pot, I., Koopmans-van Beinum, F., & Clement, C. (1998). *Influence of lack of auditory speech perception on sound productions of deaf infants.* Paper presented at the Institute of Phonetic Sciences Amsterdam.

Van Severen, L. (2012). *A large-scale longitudinal survey of consonant development in toddlers' spontaneous speech.* (Doctoral dissertation), University of Antwerp, Antwerp, Beglium.

Whalen, D., Levitt, A., & Wang, Q. (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language, 18*, 501-516.

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

ID	Gender	PTA unaided (dBHL)	Age 1st CI (y;mm.dd)	PTA with CI (dBHL)	Age fitting CI (y;mm.dd)
CI-7	F	117	0;05.05	43	0;06.04
CI-2	F	120	0;06.21	30	0;07.20
CI-9	F	103	0;08.21	32	0;09.21
CI-6	M	120	0;08.23	43	0;09.20
CI-3	F	115	0;10.00	33	0;11.20
CI-10	F	91↓117	1;01.07	43	1;02.04
CI-1	F	120	1;01.15	47	1;02.27
CI-5	M	93	1;04.27	35	1;05.27
CI-4	M	113	1;06.05	42	1;07.09
CI-8	F	112	1;07.14	52	1;09.04

Table 1: Auditory characteristics of the CI children. Legend: PTA = Pure Tone Average at

the age of 2; dBHL = decibel hearing level; HA = Hearing Aid; CI = Cochlear Implant; ↓ =

progressive hearing loss

ID	Gender	Age start (y;mm.dd)	Age end (y;mm.dd)	# utterances	# reduplicated babble	# variegated babble
CI-7	F	0;08.01	2;04.16	23	12	11
CI-9	F	1;01.24	1;05.21	13	5	8
CI-2	F	1;03.27	2;03.20	21	16	5
CI-1	F	1;04.25	2;09.27	25	18	7
CI-6	M	1;04.29	1;09.23	16	4	12
CI-5	M	1;07.24	2;01.18	20	6	14
CI-3	F	1;07.25	2;06.09	36	9	27
CI-4	M	1;11.23	2;04.12	3	2	1
CI-8	F	2;00.27	2;05.07	8	3	5
CI Total (mean; SD)	/	/	/	165 (18.33; 9.77)	75 (8.33; 5.80)	90 (10; 7.53)
NH-4	M	0;05.27	1;09.28	49	22	27
NH-1	M	0;06.03	2;00.01	94	48	46
NH-6	F	0;06.05	1;10.02	53	28	25
NH-7	M	0;07.02	1;10.02	25	8	17
NH-9	F	0;08.02	1;09.28	26	7	19
NH-5	F	0;08.30	1;11.02	29	11	18
NH-3	M	0;09.00	1;09.01	27	13	14
NH-8	F	0;11.01	1;10.26	30	7	23
NH-2	M	1;00.03	2;00.02	26	11	15
NH Total (mean; SD)	/	/	/	359 (39.88; 22.8)	155 (17.22; 13.58)	204 (22.66; 9.81)
OVERAL TOTAL	/	/	/	524 (29.11; 20.33)	230 (12.21; 11.08)	295 (15.53; 10.98)

Table 2: Recording information on the individual children. Legend: CI = Cochlear Implanted

infants; NH = Normally hearing infants; Age start: age at a the onset of babbling; Age end:

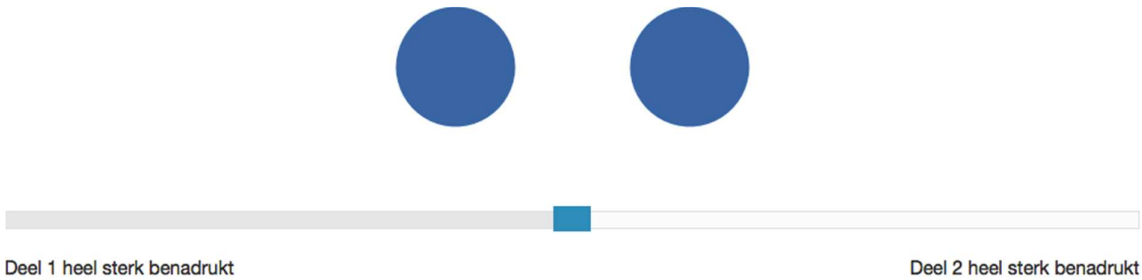age at a cumulative vocabulary of 200 words; SD = Standard Deviation

Deel 1 heel sterk benadrukt   Deel 2 heel sterk benadrukt

Figure 1a: Equally prominent syllables

Deel 1 heel sterk benadrukt   Deel 2 heel sterk benadrukt

Figure 1b: Prominent first syllable

Deel 1 heel sterk benadrukt   Deel 2 heel sterk benadrukt
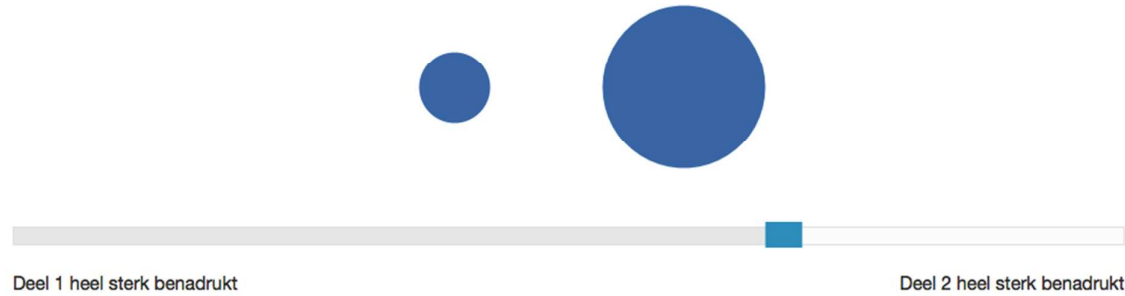
Figure 1c: Prominent second syllable

Figure 1: Print screen of three different rating positions on the sliding bar. Legend: "Deel 1/2 heel sterk benadrukt" = "Part 1/2 very strongly emphasised"
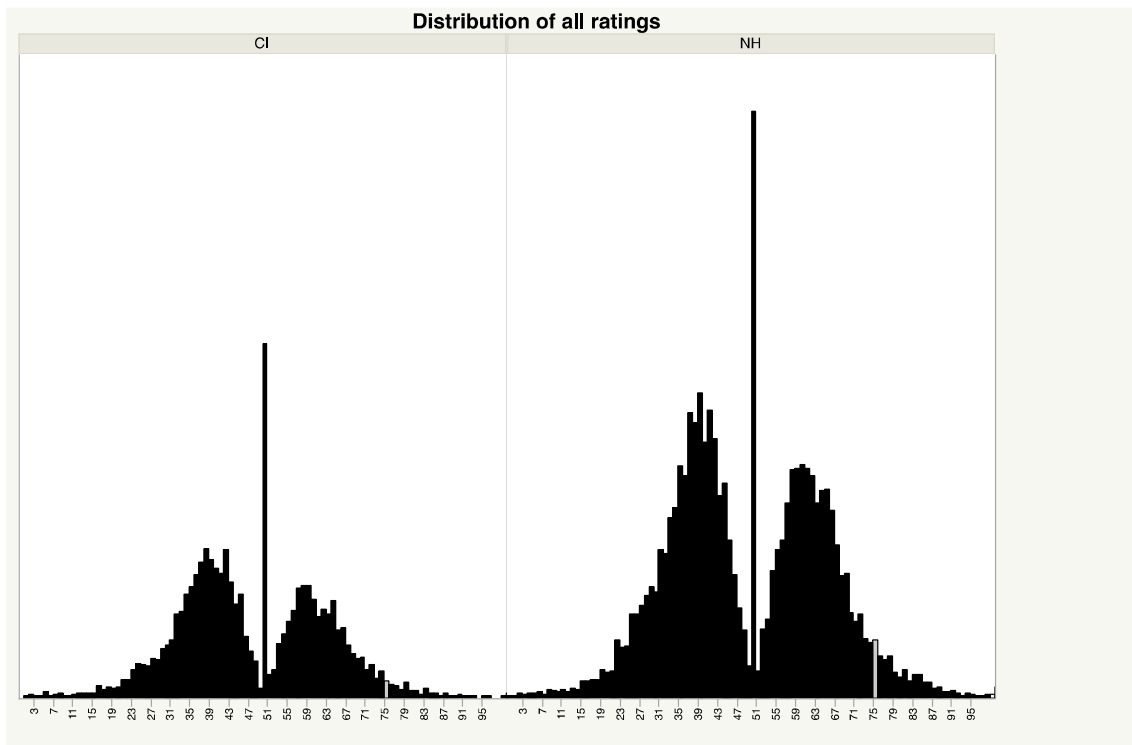
Figure 2: Distribution of all ratings per participant group.

**Supplemental material 1**

Screencast of the rating experiment (cf. mov file)

Supplemental material 2

Instructions

In this experiment you will get to hear meaningless babbles produced by infants. These babbles consist of two parts (for instance "TI TI"). These two parts can be emphasized in different gradations. For instance, the first part can be a bit, slightly or very emphasized. The same can be the case for the second part. The emphasis on part one and part two can differ or be the same.

Your task will be to indicate how the two parts relate. This will be done by means of a sliding bar that you can move to the left and the right. When you move this bar to the left, the left ball gets bigger and the right ball gets smaller. The opposite happens when you slide to the right. The size of the balls is a visual representation of the relation between the two parts of the babble: the bigger one of the two balls, the bigger the difference between the two parts. You can try out the slider below:

Deel 1 heel sterk benadrukt   Deel 2 heel sterk benadrukt
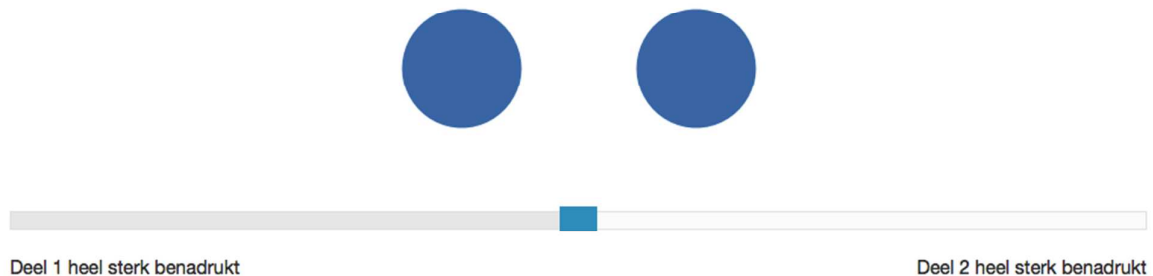
If you want to listen again to the babble you push the button below:

Herbeluister

In total you can listen to the babble three times.

Before we take of with the experiment we will show and let you hear six clear example babbles. You will hear a babble and see how the slider is moved to the right position. After these six examples you can have a try yourself on 20 babbles in which the emphasis is clear. For the first 10 trial babbles you will get a warning message if the slider is placed on the wrong side of the axis. This means that you were not able to locate the emphasis correctly. You can try again when this message occurs. If the bar is placed on the right side of the axis you can continue to the next babble.

For the last 10 trial babbles you will not get a second chance. You will have to localize the emphasis all by yourself. If you succeed in doing so, you can start with the actual experiment.

The actual experiment will look exactly the same like the trial phase.

IMPORTANT!

During the trial phase the emphasis will be clearly located on either the first or the second part of the

babble. This training is designed to familiarize you with the fact that two parts of the babble might be emphasized differently. In the babbles presented in the actual experiment the emphasis will be less clearly perceivable. Unlike the trial phase there is no right or wrong answer in the actual experiment. We expect you to follow your intuition during the experiment. What you hear is important to us.

To sum up:
- First we show you six example babbles to give you an impression of the task.
- Than the trial phase will take off. This phase consists of the judgement of 20 clear trial babbles.
  ◦ With the first 10 babbles you can only continue to the next stimulus if the bar is correctly positioned.
  ◦ With the last 10 babbles you do not get any feedback.
- If you completed the trial phase correctly you can start with the actual experiment. During the experiment there is no right or wrong answer and you just respond to what you hear. The bar on top of the experiment indicates how much progression you already made. There is no time limit so if you would like to pause during the experiment you can.

Do you have any questions left, ask the experimenter. If you are ready to take a look at the example babbles press START.

Supplemental material 3

Description of the example and trial phase

The stimuli in the example and trial phase were disyllabic reduplicative nonsense utterances, recorded by 3 female adult voices in PRAAT (Boersma & Weenink, 2014). We recorded utterances with 6 different segmental contexts, 2 consonantal contexts per vowel context (i.e. /titi/, /didi/, /koko/, /gogo/, /baba/, /wawa/). The vowels and consonants were distributed following the Frame Content Hypothesis (Davis & MacNeilage, 1990). Moreover the two consonants per vowels context were counterparts, i.e. voiced versus unvoiced (i.e. 'd-t', 'g-k'), except for 'baba' we did not use 'papa' since this is a meaningful Dutch word ('daddy'). Therefore 'wawa' was used as an alternative.

In the interactive example phase the judges got to hear and see 6 clear examples, 2 trochees, 2 iambs and 2 level utterances (see Table I below). For each example the judges saw how the sliding bar was moved in the correct position. An accompanying illustrative sentence was displayed for each example (e.g. "As you can hear the first part of the babble is more prominent than the second one"). After running through the 6 examples they could continue to the trial phase.

The purpose of the trial phase was to familiarise the judges with the rating task and see if they were competent to start with the actual experiment. In total 20 stimuli were presented to the judges. The programme automatically randomised these stimuli per judge. The task in the trial phase was the same as in the actual experiment: the adults had to indicate which part of the stimulus is more emphasised and how much. For the first 10 trial stimuli the raters could only continue to the next utterance if they had put the slider on the correct side of the axis and thus were capable to localise the stress in the utterance. If they had put the slider on the wrong side they got the message "This is not correct, please listen again". With the last 10 utterances the judges did no longer get any feedback. If less than 5 out of the 10 last utterances were judged correctly they were not allowed to start the actual experiment. In the latter case an automatic message would pop up to inform them that they could not start the experimental phase. However, none of the 30 adults had problems during the trial phase and all judges were allowed to start with the experiment.

VOICE	TROCHEE	LEVEL	IAMB
1	'ti ti	ti ti	ti 'ti
1	'ko ko	ko ko	ko 'ko
2	'ba ba	ba ba	ba 'ba
2	'ti ti	ti ti	ti 'ti
3	'ko ko	ko ko	ko 'ko
3	'ba ba	ba ba	ba 'ba
3	'di di	di di	di 'di
3	'go go	go go	go 'go
1	'wa wa	wa wa	wa 'wa
1	'di di	di di	di 'di
2	'go go	go go	go 'go
2	'wa wa	wa wa	wa 'wa

Table I: Design of the example and trial stimuli. Legend: VOICE = id of the recorded voice; ▭ =

presented in the example phase; …….. = presented in the trial phase

Supplemental material 4

Three analyses were performed in order to preserve the continuous nature of the VAS.

In the first analysis the perceived degree of prosodic modulation is the dependent variable, i.e. the more the sliding bar was moved from the midpoint of the VAS towards the extremes, the modulated the prosody. As shown in Figure 2 in the manuscript, the ratings form a multimodal distribution: a distribution on the left and the right side of the VAS and a peak at the midpoint of the VAS. The degree of prosodic modulation is represented by the ratings that move from the midpoint of the VAS towards the extremes. In order to determine the degree of modulation the scale was collapsed with midpoint = 0 and the extreme point = 50, irrespective of whether an utterance was rated at the left side or right side of the VAS. The more a rating differs from zero, the higher the degree of lexical stress. However, the residuals of the statistical model with degree of prosodic modulation as dependent variable were not normally distributed (cf. figure Ia and Ib below).
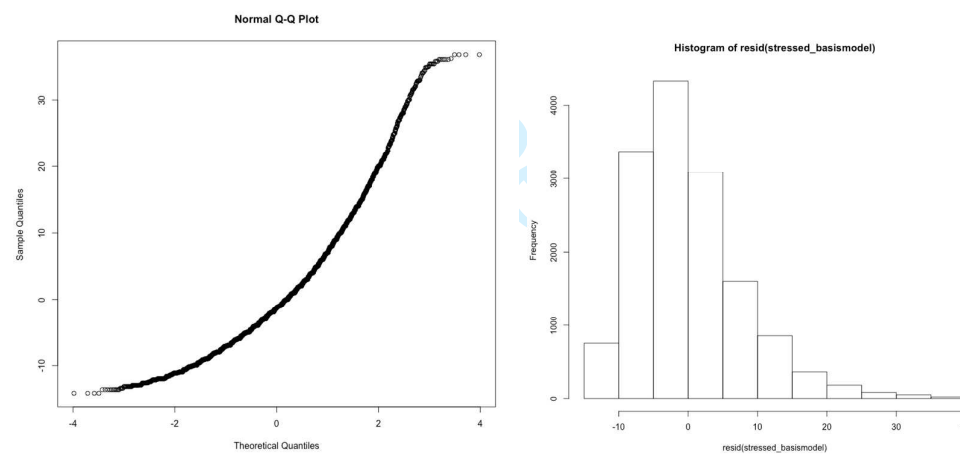

Figure Ia and Ib: distribution of the residuals of the analyses of the degree of prosodic modulation (entire dataset)

The second and the third analyses were conducted on a subset of the trochaic at the left side and iambic side. Again the dependent variable represented the degree of modulation in terms of how far the slider was removed from the midpoint. In both analyses the residuals of the statistical model were not normally distributed (cf. figure IIa, IIb, IIIa and IIIb below).
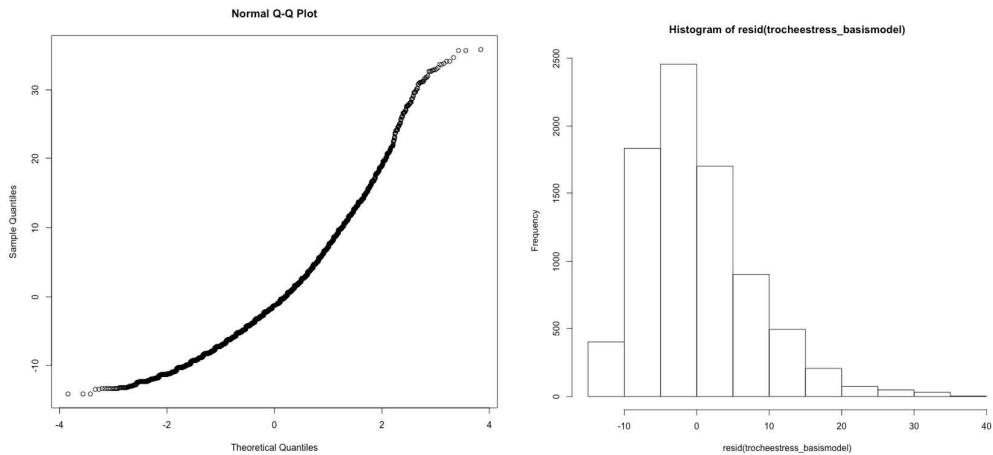
Figure IIa and IIb: distribution of the residuals of the analyses of the degree of prosodic modulation (subset of the ratings at the left/trochaic side of the VAS)
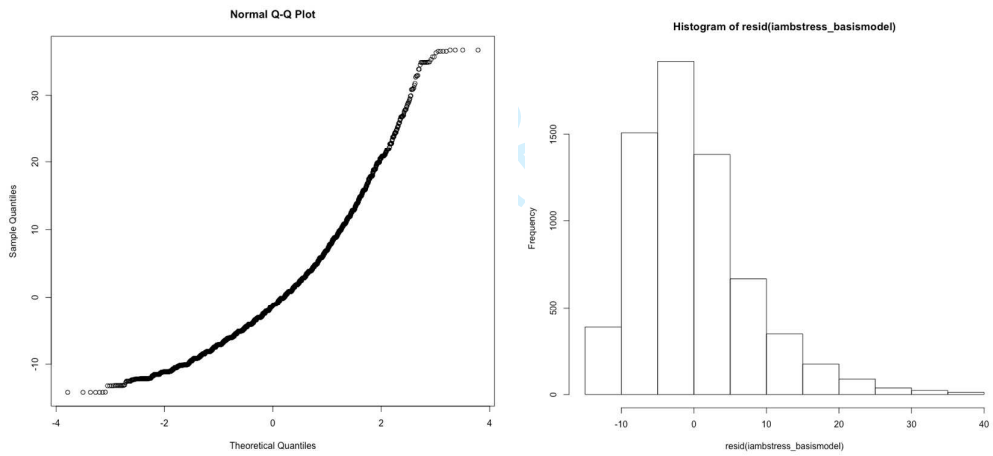

Figure IIIa and IIIb: distribution of the residuals of the analyses of the degree of prosodic modulation (subset of the ratings at the right/iambic side of the VAS)

Supplemental material 5

Analysis of the effect of participant group and utterance type on the likelihood that an utterance is rated at the midpoint of the axis (rate 50). Legend: Participant group = NH or CI; Utterance type: reduplicated or variegated babble. Significance codes: ***: p ≤ '0.001', ** ≤ '0.01', * ≤ '0.05', . ≤ '0.1'

Fixed Effects	**Estimate**	**Standard Error**	**z**	**p**
Intercept	-3.222	0.297	-10.833	<0.001 ***
Participant group (NH)	-0.401	0.188	-2.127	0.033 *
Utterance type (variegated)	-0.527	0.136	-3.878	<0.001 ***

Supplemental material 6

 Analysis of the utterances in the extreme ratings (20% most left, 20% most right, expressed in z-scores): fixed effect results expressed in logits. Legend: Participant group = NH or CI; Utterance type: reduplicated or variegated babble. Significance codes: ***: p ≤ '0.001', **: p ≤ '0.01', *: p ≤ '0.05', . : p ≤ '0.1'

Fixed Effects	**Estimate**	**Standard Error**	**z**	**p**
Intercept	-0.826	0.116	-7.133	< 0.001***
Participant group (NH)	0.255	0.129	1.981	0.048 *
Utterance type (Variegated)	0.308	0.084	3.667	< 0.001 ***
Age at onset of babbling	0.004	0.010	0.426	0.670

Supplemental material 7

Analysis of the effect of participant group (NH or CI) on the repetitions of the stimuli. Legend:

Participant group = NH or CI. Significance codes: ***: $p \leq$ '0.001', ** $\leq$ '0.01', * $\leq$ '0.05', . $\leq$ '0.1'

Fixed Effects	Estimate	Standard Error	t	p
Intercept	0.423	0.046	9.156	< 0.001***
Participant group (NH)	-0.058	0.017	-3.349	0.004 **
20% Extreme	-0.126	0.009	-14.613	< 0.001***

Supplemental material 8

Analysis of the effect of participant group on the likelihood that an utterance is rated as a trochee.

Legend: Participant group = NH or CI. Significance codes: ***: $p \leq$ '0.001', ** $\leq$ '0.01', * $\leq$ '0.05', . $\leq$ '0.1'

Fixed Effects	Estimate	Standard Error	z	p
Intercept	0.157	0.268	0.586	0.558
Participant group (NH)	- 0.041	0.345	-0.120	0.904

Supplemental material 9

Analysis of the effect of participant group and utterance type on the likelihood that an utterance is rated as an iamb. Legend: Participant group = NH or CI; Utterance type: reduplicated or variegated babble. Significance codes: ***: $p \leq$ '0.001', ** $\leq$ '0.01', * $\leq$ '0.05', . $\leq$ '0.1'

Fixed Effects	**Estimate**	**Standard Error**	**Z**	**p**
Intercept	-0.929	0.320	-2.905	0.004 **
Participant group (NH)	0.094	0.389	0.243	0.808
Utterance type (variegated)	0.401	0.192	2.088	0.037 *