Listeners' perception of lexical stress in the first words of infants with cochlear implants and normally hearing infants

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

[a] Department of Linguistics, CLiPS Computational Linguistics and Psycholinguistics research centre, University of Antwerp, Prinstraat 13, Antwerp, Belgium.

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

Corresponding author:

Ilke De Clerck, Department of Linguistics, CLiPS, University of Antwerp, Prinsstraat 13, 2000 Antwerp, Belgium, Tel. +3232655239, E-mail ilke.declerck@uantwerpen.be

E-mail addresses authors:

Jo Verhoeven: jo.verhoeven.1@city.ac.uk

San Gillis: gillissan@gmail.com

Michèle Pettinato: michèle.pettinato@uantwerpen.be

Steven Gillis: steven.gillis@uantwerpen.be

Abstract

Normally hearing (NH) infants are able to produce lexical stress in their first words, but congenitally hearing-impaired children with cochlear implants (CI) may find this more challenging, given the limited transmission of spectro-temporal information by the implant. Acoustic research has shown that the acoustic cues to stress in the first words of Dutch-acquiring CI infants are less pronounced (Pettinato et al. 2017). The present study investigates how listeners perceive lexical stress in the first words of CI and NH infants. Two research questions are addressed: (1) How successful are CI and NH children in implementing the prosodic cues to prominence? (2) Is the degree of stress in CI and NH words perceived to be similar?

The stimuli used in this study are disyllabic words (n = 1089) produced by 9 infants with CI and 9 NH infants acquiring Dutch. The words were presented to adult listeners in a listening experiment, in which they assessed the stress pattern on a continuous visual analogue scale (VAS) which expresses to what extent syllables are perceived as stressed.

The results show that listeners perceive typical word stress production in the first words of infants with CI. The words of CI and NH infants were rated in agreement with the target stress pattern as often, and trochaic words were rated more frequently as such than iambic words. Listeners more frequently perceive unstressed syllables in the first words of infants with CI. However, for the words that are perceived to be clearly stressed, the degree of word stress is comparable in the two groups, and both infant groups are perceived to produce more contrast between stressed and unstressed syllables in trochees than in iambs. It is concluded that that acoustic differences between CI and NH infants' stress production are not necessarily perceptually salient.

Highlights

- The present study investigates how listeners judge lexical stress in the first words of CI and NH infants on a visual analogue scale.
- Words of infants with CI are as often rated in agreement with the target stress pattern as the words of NH infants.
- Listeners more frequently perceive no stressed syllable in the first words of infants with CI as compared to the first words of NH infants.
- It is concluded that acoustic differences between CI and NH infants' stress production are not necessarily perceptually salient.

1 Introduction

This study investigates the perception of lexical stress in the early words of Dutch congenitally hearing-impaired infants with a Cochlear Implant (henceforth, CI) and their normally hearing (NH) peers. Lexical stress refers to the phenomenon that a syllable in a word is perceptually more salient than others. Acoustically stressed syllables are characterised by an increase in syllable duration and intensity, and substantial change in fundamental frequency (F0). Listeners perceive these stressed syllables as longer, louder and higher pitched than unstressed syllables (Lieberman, 1960). From a phonological point of view, Dutch is a predominantly trochaic language as the majority of disyllabic words carry stress on the first syllable (e.g. /'ʌuto/, 'car'). However iambic words, i.e. disyllables with a stressed second syllable, also occur (e.g. /bɑ'nan/, 'banana'). About 69.73% of Dutch disyllables are trochaic and 30.27% are iambic (Hide, 2013).

NH infants are able to perceive the prosodic pattern of their ambient language from early on in development (Querleu, Renard, Versyp, Paris-Delrue, & Crepin, 1988; Sansavini, Bertoncini, & Giovanelli, 1997), and research has shown that they also have command over the prosodic cues to stress in their own production (Davis, MacNeilage, Matyear, & Powell, 2000; De Clerck, Pettinato, Verhoeven, & Gillis, 2017; DePaolis, Vihman, & Kunnari, 2008). However, for children with severe-to-profound hearing loss the adequate perception of prosody is impaired, as they only have limited access to prosodic information in the speech signal. Consequently their own production of prosody is impaired (Clement, 2004; Kent, Osberger, Netsell, & Hustedde, 1987; van den Dikkenberg-Pot, Koopmans-van Beinum, & Clement, 1998).

For infants with severe-to-profound sensorineural hearing loss, cochlear implantation is a common technique to (partially) restore hearing. Cochlear implantation early in development leads to substantially better language outcomes as compared to hearing-impaired infants with acoustic hearing aids (Dettman et al., 2016; Levine, Strother-Garcia, Golinkoff, & Hirsh-Pasek, 2016; Tomblin, Barker, Spencer, Zhang, & Gantz, 2005). However, a CI device remains limited in transmitting spectro-temporal information, causing inadequate perception of F0 (Green, Faulkner, & Rosen, 2004; Moore, 2003; O'Halpin, 2010) and intensity (Drennan & Rubinstein, 2008; Meister, Landwehr, Pyschny, Wagner, & Walger, 2011; Moore, 2003): these are the two most important cues to prosodic prominence. Given the suboptimal transmission of prosodic information by CI devices, it can be expected that stress production may well be affected in infants with CI.

Typically developing infants are sensitive to prosody from birth (Sansavini et al., 1997), and even as early as the last trimester of the pregnancy (Querleu et al., 1988). Infants are already sensitive to the prosodic pattern of their native language during the first year of life (Friederici, Friedrich, & Christophe, 2007; Jusczyk, Cutler, & Redanz, 1993). This early perceptual sensitivity has a beneficial effect on prosody production from early on in development. Research has shown that the production of ambient prosody becomes well established from word-use onwards (Davis et al., 2000; De Clerck et al., 2017; DePaolis et al., 2008). Davis et al. (2001) investigated prosodic prominence in the canonical babble of English NH infants, and found that listeners only judged half of the babble to have a clear prominent syllable. However, in the utterances with a prominent syllable, the infants did manipulate the prominence cues (i.e. F0, intensity and duration) to the same extent as a control group of adult speakers. This suggests an emerging ability to use the acoustic cues to prominence in prelexical utterances, but without a clear bias towards the predominant English stress pattern (i.e. trochaic pattern). DePaolis et al. (2008) suggest that the production of the predominant stress pattern becomes well-established when the infants produce more words than babble, i.e. around the age of 18 months. They conducted a cross-linguistic study on two iambic languages (French and Welsh) and two trochaic languages (English and Finnish). Disyllabic utterances produced at the 4-word-point (i.e. the recording in which a cumulative vocabulary of 4 words was registered) were selected. Acoustic measurements showed no differences in the number of trochees and iambs, meaning that the predominant stress pattern of every language in the study was not yet apparent in babble. This seems to suggest that infants need longer linguistic experience to enhance the production of the predominant stress pattern of the ambient language.

In line with these studies, a recent study on the speech of NH infants has shown that word-use boosts the production of prosodic prominence (De Clerck et al., 2017). Disyllabic babble and words were selected from monthly recordings of nine typically developing infants acquiring Dutch. The prosodic differentiation of these utterances was investigated by measuring F0, intensity and duration in both syllables. The results showed firstly that the prosodic differentiation was clearer in first words compared to babble. Secondly, the predominant Dutch stress pattern (i.e. the trochaic pattern) was already apparent in babble, as evidenced by a higher F0 and (to a lesser extent) intensity of the first syllable. This pattern became clearer in the children's first words. Thirdly, De Clerck et al. (2017) showed that the prosodic differentiation did not gradually improve as the infants' vocabulary increased, but that word-use itself boosted the ability to produce the phonetic features of the ambient language. From this, it can be concluded that NH infants are able to reliably implement prosodic prominence in their first words.

For congenitally hearing-impaired children with CI, stress production in their first words may not be as straightforward as for their NH peers. First of all, they have only had very limited exposure to the prosodic characteristics of the ambient language before receiving their implant. Even if they are implanted at a very early age, CI children miss out on several important months of speech experience during which NH infants already appear to become familiar with the prosodic characteristics of the ambient language. In addition, CI devices are limited in transmitting spectro-temporal information, so that CI children may not be able to access information on fundamental frequency and intensity required for normal prosody production. Most research has shown that prosody perception is suboptimal in CI users (Most & Peled, 2007; O'Halpin, 2010; Titterington, Henry, Kramer, Toner, & Stevenson, 2006). However, recent studies have also indicated that CI devices transmit sufficient acoustic information to allow for stress pattern recognition (Segal, Houston, & Kishon-Rabin, 2016; Vavatzanidis, Mürbe, Friederici, & Hahne, 2016). Some of these studies also show that CI infants can detect different stress patterns, although their discrimination abilities remain poorer than those of NH infants.

Given the more restricted exposure and the reduced perception of prosody, it may well be the case that the production of prosody is also affected in infants with CI. Although research on prosody production shortly after implantation is rare, it has been shown that school-aged children with CI experience problems with prominence production (Carter, Dillon, & Pisoni, 2002; Hide, 2013; Lenden

& Flipsen, 2007). Stress production by 3- to 6-year-old CI users (mean age of implantation 2 years, 4 months; mean hearing age 2 years, 8 months) was described as sounding ‘excessive, equal or misplaced’ instead of sounding like the target stress pattern (Lenden & Flipsen, 2007). Problems with prominence production were also found in non-word repetition tasks: Carter et al. (2002) showed that 8- to 10-year-old English children with CI (mean age of implantation 3 years, 3 months; mean hearing age 5 years, 3 months) correctly imitated the stress pattern of only 61% of non-words. Hide (2013) showed that 6- to 9-year-old CI children acquiring Dutch 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% correct imitations for trochaic and 97% for iambic non-words). Moreover, Hide (2013) acoustically measured the non-words that were unanimously judged as having stress on the first or the second syllable. The measurements showed less acoustic differentiation (both regarding pitch excursion and pitch excursion duration) between the syllables of CI utterances in comparison to NH utterances.

The above-mentioned studies show a discrepancy between stress production of school-aged NH and CI children. Since NH infants already produce prosodic prominence in their first words, the question arises whether a discrepancy between CI and NH children’s prosody can be detected earlier in development. A recent acoustic study on the infants included in the present study has shown that children with CI do experience problems with producing prosodic differentiation between the syllables of their disyllabic babble and first words (Pettinato, De Clerck, Verhoeven, & Gillis, 2017). F0, intensity and duration were measured in the disyllabic babble and words of infants with CI and NH infants. CI infants produced less differentiated F0-values from babbling onwards, and this discrepancy between CI and NH infants became even larger in their first word productions. The F0 values of the NH infants showed a shift towards the predominant trochaic pattern (i.e. stress on the first syllable) in word productions, whereas this was not the case in the words of CI infants. A similar but smaller tendency was found for intensity measurements. CI infants also showed a smaller durational contrast between syllables of lexical utterances, which creates a less differentiated stress pattern than their NH peers.

The present study investigates the perception of lexical stress in disyllabic CI and NH words. Although the acoustic study of Pettinato et al. (2017) showed deviant prominence production in the first words of CI children, the exclusive focus on measuring individual prosodic cues is too restricted. Firstly, subtle acoustic differences reflected in the measurements of acoustic dimensions are not necessarily perceptually salient to listeners. A perceptual study is necessary to put acoustic prosodic differences into perspective. ‘t Hart, Collier, and Cohen (1990) have clearly shown that perceptual research on intonation is vital to accurately assess the perceptual relevance of acoustic differences. Research on segmental development shows a similar discrepancy between acoustic measurements of speech productions and the perception by listeners. Acoustic differences between speech sound categories evident from measurements are not necessarily perceived by listeners, i.e. so called covert contrasts (Li, Edwards, & Beckman, 2009; Scobbie, Gibbon, Hardcastle, & Fletcher, 1996). The same may be the case for the prosodic discrepancy between CI and NH words: the acoustic differences between groups measured by Pettinato et al. (2017) may not be perceived as relevant by adult listeners. This implies that the prosody production of CI infants is not so divergent that it impacts their intelligibility or that listeners would perceive their speech as impaired.

The second reason to use a perceptual experiment is that the acoustic measurements of single parameters may not give a conclusive image of the stress pattern in a word. Prosodic cues maintain trade-off relations in speech (Lieberman, 1960), which means that not all three prosodic cues may contribute equally to the perception of syllable prominence. That is, when intensity is higher in the first syllable and pitch and duration are higher in the second syllable, these measurements do not necessarily mean that the word is perceived as a trochee (or an iamb, for that matter). Therefore, perceptual judgements are necessary to complement acoustic measurements, since listeners make prominence judgements by taking into account all cues at the same time (Flege & Bohn, 1989; Fry, 1958).

The aim of the present study is to investigate the perception of lexical stress in the first words of CI and NH infants. This is motivated by the observation that the acquisition of prominence may be more challenging for infants with CI, given the limited transmission of prosodically relevant information by the implant. Previous studies have shown impaired prosody production in the speech of school-aged CI children (Carter et al., 2002; Hide, 2013; Lenden & Flipsen, 2007) and in acoustic measurements in first words (De Clerck et al., 2017). The present paper investigates whether the smaller acoustic differentiation in CI words is perceptually salient, since early detection of such differences may lead to earlier clinical interventions. Hunter, Kronenberger, Castellanos, and Pisoni (2017) have recently pointed out the importance of the early detection of speech deficits in infants with CI. They show that speech perception and language skills measured 6 and 18 months post-implantation are predictors of long-term language outcomes (measured by open word set recognition) and working memory outcomes in CI users. The importance of early detection of speech deficits is a motivation to investigate prosody production in the early speech of CI users.

The present study was conceived around two research questions:

Research question 1: How successful are CI and NH children in implementing the prosodic cues to prominence? This addresses the issue of whether both groups of children are able to successfully mark the stressed syllable for prominence. As a result of deprived perception of the prosodic cues to prominence, it is hypothesised that CI children are not as successful in marking stress as their NH peers. This is reflected perceptually by the fact that CI words are not perceived as often as having the stress pattern of the target word.

Research question 2: Do listeners perceive the same degree of lexical stress in the first words of CI and NH infants? Since CI infants have less fine-grained prosody perception, and since previous acoustic measurements have shown less differentiated stress patterns, it is expected that adult listeners perceive the stresses in CI words to be weaker than in NH words.

2 Method

2.1 Participants

For the purpose of this study, 9 infants with a cochlear implant and 9 normally hearing children were included from the CLiPS Child Language Corpus, a collection of longitudinally collected video recordings and their transcriptions (Molemans, 2011; Schauwers, 2006; Van den Berg, 2012; Van Severen, 2012). The use of these data has been approved by the Ethics Committee for the Social sciences and Humanities (EA SHW) of the University of Antwerp (SHW_15_37). The same infants with CI from the study of Pettinato et al. (2017) and the same NH infants from the studies of De Clerck et al. (2017) and Pettinato et al. (2017) are included in the present study. The following paragraphs summarise the applied recruitment procedure and participant characteristics.

The children with CI had been recruited from an Academic ENT Unit in Antwerp/Belgium. These children 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 health or developmental problems were reported. The children had been implanted before the age of two, ranging from 5 to 19 months (*M* = 12 months; *SD* = 5 months). The average unaided hearing loss was 113 dBHL in the better ear. Before implantation, the range of the Pure Tone Averages (PTA) was 93-120 dBHL (*M* = 113 dBHL; *SD* = 9 dBHL). One year after implantation, the PTA 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. All children had been raised in monolingual homes acquiring Belgian Dutch (Verhoeven, 2005). The auditory characteristics can be found in Table 1.

INSERT TABLE 1 ABOUT HERE

As a control group for this study, 9 NH children were randomly selected from the CLiPS Child Language Corpus. These children had been recruited from day-care centres, families known by the researchers and by advertisements. The typical development of these children was established on the basis of a parental report and a checklist of the attainment of communicative and motor milestones based on a questionnaire developed by ‘Kind en Gezin’, the Flemish infant welfare centre (Molemans, 2011). Normal language development was verified by means of the Dutch version of the CDI, i.e. N-

CDI (Zink & Lejaegere, 2001) administered at ages (years; months) 1;0, 1;6 and 2;0. All children were raised in monolingual homes acquiring Belgian Dutch (Verhoeven, 2005).

Both groups of children have been videotaped on a monthly basis. The recordings of the NH children started at 6 months and lasted up to 24 months. The CI children were recorded from the month they received their implant up to 30 months post implantation. Only a subset of the recordings included in the acoustic studies of Pettinato et al. (2017) and De Clerck et al. (2017) were included in the present study. Since the aim of the present study is to investigate lexical stress in first words, only recordings were selected in which the children spoke their first identifiable words. The range of monthly recordings started from the recording in which the onset of word use was registered and ended with the recording in which a cumulative vocabulary of 200 words was attested. This cut-off point was motivated by the fact that the selected range should provide enough early words for every child included in this study. The mean age of the CI children at the start of the recordings (i.e. the onset of word use) was 22 months (*SD* = 3 months). The mean age at the cut-off point was 30 months (*SD* = 2 months). The mean age of the NH children at the start of the recordings (i.e. the onset of word use) was 15 months (*SD* = 2 months). The mean age at the cut-off point was 23 months (*SD* = 2 months). The ages of the individual children at the time of recording are given in Table 2.

INSERT TABLE 2 ABOUT HERE

2.2 Perceptual experiment

The aim of this study was to investigate whether there is a perceptual difference in stress in the early words of CI and NH children. For this purpose, a listening experiment was carried out in which adult participants assessed the placement and the degree of lexical stress in disyllables.

2.2.1 Adult listeners.

Ninety adults participated as listeners in the perceptual experiment (Mean age: 23.2, SD: 6.2). They were native speakers of Dutch. The listeners were not informed about the purpose of the study. Moreover, they were not told that the recordings were taken from two groups of children. During recruitment, the listeners filled out a questionnaire (Supplemental material 1). The listeners had to indicate how frequently they came in touch with infants. The listeners in the present experiment were not particularly familiar with child speech. Moreover, none of the listeners reported problems with hearing or other health problems.

2.2.2 Stimulus selection.

The selected monthly recordings in CLiPS Child Language Corpus were used for stimulus selection. From these recordings, disyllabic early words were extracted using a set of selection criteria. All details about the data and the selection criteria can be found in Pettinato et al. (2017) and in De Clerck et al. (2017) and are summarised briefly in this section.

The child utterances in the recordings were phonemically transcribed. On the basis of this transcription, the adult target forms of the produced words were determined following the guidelines of Vihman and McCune (1994). Based on the phonemic transcription, the stress pattern of the target word was retrieved from the pronounciation database Fonilex, which is a database of the most frequent word forms in spoken Dutch (Mertens, 2001).

Only disyllabic utterances were selected for the present experiment. An utterance was considered to be a disyllable when it consisted of two vocalic phases minimally separated by a clear consonantal phase. Additional consonants flanking the vocalic sections were allowed. Since the present study is a perceptual experiment and the studies from Pettinato et al. (2017) and De Clerck et al. (2017) were acoustical studies, only the words that were suitable for a perceptual study were included in the present experiment. The sound quality of the selected utterances had to be high enough: disyllables were only 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 1089 disyllabic early words (529 CI words, 560 NH words) were selected as stimuli for the perceptual experiment (see Table 2).

The children's word productions were disyllables, but this did not necessarily mean that the attempted adult target word was a disyllable, e.g. the adult target 'banana' has three syllables but if pronounced as 'nana' it met the criterion for a disyllabic utterance in this study. In total, 330 different target words were produced, of which 258 were disyllabic. Supplemental material 2 provides an overview of the attempted target words produced by the two groups of children (i.e. different types, total number of types and tokens).

The 1089 selected stimuli were presented to adult listeners in a perceptual experiment. Since there were too many stimuli (1089) to be rated by one single listener, the workload of the experiment was reduced by dividing the 1089 stimuli into three subsets, each containing 363 stimuli. Each subset was presented to 30 participants, so that that every word was rated 30 times. Every subset contained about the same number of stimuli from CI and NH infants. The experimental interface automatically randomised the stimuli within every subset, in order to control for confounding order effects.

2.2.3 Experimental design.

In the present study, listeners indicated their perceptual ratings on a Visual Analogue Scale (henceforth, VAS). A VAS is a psychometric measurement tool consisting of a line with two opposite characteristics of a stimulus at the extremes of the scale. In this case the extremes of the VAS represent "a very prominent first syllable" (i.e. trochee) versus "a very prominent second syllable" (i.e. iamb). Listeners have been shown to be able to indicate subtle phonetic differences on a VAS (Julien & Munson, 2012; McAllister Byun, Harel, Halpin, & Szeredi, 2016; Munson, Schellinger, & Carlson, 2012). Therefore, a VAS is considered to be a valid tool to map subtle phonetic differences in early speech.

A screen shot of the VAS interface is given in Supplemental Material 3. Three different rating positions are illustrated in Figure 1. The listeners heard a stimulus and had to assess: (a) the stress pattern, by moving the slider to the left (i.e. a prominent first syllable), to the right (i.e. a prominent second syllable), or leave it in the middle of the scale (i.e. no clearly prominent syllable), and (b) the degree of prominence, by moving the slider from the midpoint towards the extremes. Two circles above the sliding bar served as a visualisation of the two syllables of each stimulus. The initial position of the slider was at the midpoint of the VAS (Figure1a). When sliding to the left, the left circle became larger while the right one became smaller, indicating a more prominent first syllable (Figure1b). 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 (Figure1c). The position of the slider moved along a scale ranging from 0 (extreme left) to 100 (extreme right). Thus, a rating between 0 to 49 means more prominence on the first syllable (the lower the number, the more stress on the first syllable). Fifty indicates equal stress on both syllables and a rating between 51 to 100 indicates more stress on the second syllable (the higher the number, the more stress on the second syllable).

Before the start of the experiment, the listeners received information about the task and received a short training. In the information phase, the goals of the experiment were explained in a section with instructions (see Supplementary Material 4). In the training phase, the participants were required to indicate the stress in some clear examples.

INSERT FIGURE 1 ABOUT HERE

In the actual experiment, a set of 363 words was presented in random order to the listener. The experiment ran on an iMac (type 21.5 inch Late 2013, 2.9 GHz Intel Core i5) in a quiet room. The stimuli were presented via SONY MDR-1R headphones. Every stimulus was preceded and followed by one second of silence. The total duration of the experiment was approximately 60 minutes.

For every stimulus, the following information was registered for statistical analysis: the unique number of every stimulus ('utterance identity': 1-1089), the identity of the infant ('child identity': 1-18), the hearing status of the child ('participant group': NH or CI), the blinded identity of the adult judge ('listener identity': 1-90), the VAS-score given to the stimulus ('rate': 0-100), the attempted adult target word produced by the infant ('target word'), the stress pattern of the attempted adult target word ('target stress pattern': trochee or iamb) and the identifiability of the word.

2.2.4 Procedure for determining the identifiability of the words.

Since the stimuli are children's meaningful word productions, the adult listeners participating in the experiment should in principle be able to understand them, and their perception may be influenced by the target stress pattern. For instance, when a listener recognises the trochaic word 'auto' (i.e. 'car'), he may be more inclined to rate it as a trochee even though the child's rendition is not clear. In such cases, a trochaic bias in the ratings is not necessarily or entirely attributable to the infant's prominence production but to the target stress pattern. However, since the stimuli in the experiment are presented out of context without the support of video images and outside their conversational context, it may well be that the listeners do not recognise some words. Consequently, some stimuli may be recognised while others remain unidentifiable. In order to examine this factor, the identifiability of the stimuli was determined and used as a predictor in the statistical analyses.

In order to determine the identifiability of the stimuli, they were presented to four listeners. All these words were presented in isolation without any conversational context. The four adult listeners were instructed to phonemically transcribe the stimuli. The aim was to determine whether it was be possible to recognise the words in isolation. For each stimulus, the number of times it was identified was registered (i.e. 0 to 4). About half of the stimuli were identified by none of the four listeners (CI: 47% of the words, NH: 51% of the words). 23% of the CI words was identified by all 4 listeners, for the NH group this was 19%. Mean identifiability per stimulus was 1.5 (SD = 0.30) for the CI group and 1.4 (SD = 0.30) for the NH group. This means that the target words were difficult to identify without conversational context.

2.2.5 Statistical approach

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 which are structured hierarchically. Moreover, GLMM is robust to missing data and different numbers of word productions per child. These models consist of a random and fixed part. The random part takes into account the variation caused by the random effects: in the present study there is variation between individual infants (n = 18), individual listeners (n = 90) and different stimuli nested (n = 1089) in different target words (n = 330). The fixed part consists of the predicting or independent variables: participant group, target stress pattern, cumulative vocabulary and identifiability of infants' word productions.

The statistical models were constructed iteratively: first the random effects were added one by one, then the fixed effects were added one at the time and finally the interaction effects were added. After adding a random or fixed effect, the resulting model was compared to the previous model by means of a likelihood ratio test. Variables were added to the model if they improved the fit of the model at the time they were entered. Previously added variables were not removed if they became non-significant after other variables were added in a subsequent step. The best-fitting model is reported in the results section.

The same procedure was followed in all analyses. First of all, the random effects were added in the following order: participant identity, target word and stimulus. Secondly, the fixed effects were added. First the most important predictive variables were added one at a time (i.e. participant group and stress pattern of target word), next the other predictive variables were added one at a time (i.e. identifiability and cumulative vocabulary). In a next step, the following interactions between variables were tested (in the respective order): participant group x stress pattern of the target word, participant group x identifiability, participant group x cumulative vocabulary. The interaction effects between 'participant group' and the other fixed effects were tested to see whether the same effects are found in both infant groups. The best fitting models are reported in the results section.

The first research question, *How successful are CI and NH children in implementing the prosodic cues to prominence?* focuses on the relation between the ratings and the target stress pattern of the words. This statistical analysis examines whether the ratings on the VAS are in agreement with the stress pattern of the targets. For this purpose, the analysis was done on a subset that only included the words with a disyllabic target word (CI = 443 disyllabic target words, NH = 488 disyllabic target words. See supplemental material 2). The target stress pattern is either trochaic or iambic. The target stress pattern was then compared to the actual ratings (i.e. rated at the trochaic or iambic side of the axis, or at the midpoint of the VAS). For every word in this subset, a proportion of the agreement between the ratings and the target stress pattern was calculated: the number of ratings that agreed with the adult target was divided by the total number of ratings (i.e. 30), resulting in a proportion between 0 (i.e. no agreement between the stress pattern in the child's production and the target stress pattern) and 1 (i.e. all ratings were the same as the target stress pattern). The dependent variable in the first analysis is the proportion of correct ratings of the stress pattern. The random and fixed effects were added as described in the previous paragraph.

The second research questio*n, Do listeners perceive the same degree of word stress in the first words of CI and NH infants?* is investigated by two statistical analyses. The first analysis examined the degree of word stress. In the experiment, listeners indicated the degree of stress on the VAS by moving the slider from the midpoint towards the left or right extreme. The position of the slider on the VAS thus represents the degree of prosodic contrast between the two syllables of a word. The focus of the first analysis is twofold: (1) Did the listeners move the slider more towards the extremes of the VAS for one of the two groups of children or is there no significant difference? and (2) Are words that are perceived as trochaic rated with as much stress as words that are perceived as iambic, or is there no significant difference? The dependent variable in this analysis is the degree of word stress. The random and fixed effects were added in the same order as in the first analysis. The only difference is that, in the random part, stimuli were nested in target words. In the fixed effects, rating location was added to the model. Cumulative vocabulary was added consistently as the final predictive variable in all models.

In a second analysis, the absence of lexical stress was investigated: are the stimuli of one of the two groups rated more frequently at the midpoint of the VAS? The dependent variable in this final analysis is the binomial variable rated at the midpoint or not. This binomial dependent variable was analysed by means of a Logistic Regression Analysis in the form of GLMM. The random and fixed effects were added in the same order as in the first analysis. The only difference is that in the random part stimuli were nested in target words.

INSERT FIGURE 2 ABOUT HERE

3 RESULTS

3.1 Analysis of the stress pattern

Are the words of CI and NH infants rated with equal frequency in agreement with the stress pattern of the target word? Table 3 gives an overview of the occurrence of the stress patterns of the attempted target words, i.e. the proportion of trochees and iambs in the subset of disyllabic target words. The majority of the stimuli has a trochaic target, which reflects the predominant stress pattern in Dutch disyllabic words. Secondly, Table 3 also shows the proportion of the ratings that are in agreement with the target stress pattern: in what proportion of the case do the listeners actually hear the target stress pattern and indicate the appropriate region on the VAS? A score of 0 means that no rating is in agreement with the target stress pattern and a score of 1 means that all ratings are in agreement with the target stress pattern. Of all utterances, trochaic targets are rated correctly most frequently, and there does not seem to be a large difference between CI and NH infants.

INSERT TABLE 3 ABOUT HERE

The best fitting model is reported in Table 4. The random part of the best fitting model controlled for the variance explained by participant identity and the target word. The fixed effects required in the best fitting model were participant group, the stress pattern of the target word, identifiability of the stimulus and cumulative vocabulary. The intercept of the model was estimated at 0.238 ($p < 0.001$). The results show no significant group effect, meaning that NH words are not rated more in agreement with the stress pattern of the target word ($p = 0.366$). The target stress pattern does significantly influence the proportion of correctly rated stress patterns: the proportion of correctly rated stress patterns is significantly higher for trochaic target words than for iambic target words ($p < 0.001$). The identifiability of a word positively impacts the agreement between ratings and stress patterns: the more an utterance is identified, the more it will be rated in agreement with the target stress pattern ($p < 0.001$). There is also a small impact of increasing cumulative vocabulary on the proportion of correctly rated stress patterns: the larger the cumulative vocabulary of the children, the more their utterances will be rated in agreement with the target stress pattern ($p = 0.030$). This is the case for both CI and NH words, but the effect is slightly larger for the NH words ($p < 0.001$). To sum up, this analysis mirrors the descriptive statistics that are displayed in Table 3: there is no significant difference between the two groups regarding the agreement between listeners' ratings and stress patterns, and in both groups, utterances with a trochaic target are rated most frequently in agreement with the target stress pattern.

INSERT TABLE 4 ABOUT HERE

3.2 Analysis of the degree of word stress

Do listeners perceive the same degree of word stress in the first words of CI infants and NH infants? Listener's perception of the degree of stress is investigated by looking at (a) the distance of the slider from the midpoint of the VAS, and (b) the likelihood that the slider is at the midpoint of the VAS.

3.2.1 Degree of word stress

In the first analysis, the perceived degree of lexical stress is the dependent variable, i.e. the more the sliding bar was moved from the midpoint of the VAS towards the extremes, the more prosodic differentiation was perceived. As shown in Figure 2, 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 lexical stress is represented by the ratings that move from the midpoint of the VAS towards the extremes. In order to determine the degree of stress, the scale was collapsed with midpoint = 0 and the extreme point = 50 (see Figure 3), 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. Although this distribution is not normal, a multilevel analysis is permitted since the residuals of the estimated MLM are approximately normally distributed.

INSERT FIGURE 2 and 3 ABOUT HERE

The best fitting model is reported in Table 5. The random part controlled for the variance of participant identity, listener's identity and the stimuli nested in the different target words. The required fixed effects were participant group, the identifiability of the word, and 'side of the rating' i.e. whether the rating was located at the trochaic or iambic side of the VAS. The intercept of the model was estimated at 14.353 ($p < 0.001$), meaning that the degree of stress of an average CI word is situated around this value. The results show more stress in NH words, but this effect is not significant ($p = 0.111$). The identifiability of the words has a negative effect on the degree of stress ($p = 0.011$), meaning that a more identifiable word is rated with less stress. Moreover, words that are rated on the trochaic side of the VAS are perceived to be more stressed ($p < 0.001$), and this is the case for both groups of infants. Adding the interaction between infant group and stress pattern did not significantly improve the fit of the model. This means that for both groups, utterances that are rated at the trochaic side of the VAS are also perceived to be more stressed.

To sum up, the main finding of this analysis is that listeners perceive no gradual difference in stress production of CI and NH infants, and both groups are perceived to produce a higher degree of stress when they stress the first syllable than when they stress the second syllable.

INSERT TABLE 5 ABOUT HERE

3.2.2 Midpoint of VAS

In a second analysis, the probability of utterances rated at the midpoint of the VAS (score: 50) was investigated. The midpoint represents the assessment that the 2 syllables are perceived as equally stressed. Figure 2 shows the distribution of all ratings on the VAS. A peak is centred on the midpoint (i.e. score 50) of the VAS. This was the default position of the sliding bar. This point on the VAS is considered to represent the words with no clear prominent syllable. The ratings were recoded into a dummy variable: rated at the midpoint versus rated anywhere else on the rating scale.

This binomial dependent variable was analysed by means of a Logistic Regression Analysis in the form of GLMM. The aim of this analysis is to investigate whether utterances of infants with CI and NH infants were equally likely to be rated at the midpoint. The results are reported in Table 6. These values are expressed in logits, but to facilitate the interpretation they are converted to probabilities.

INSERT TABLE 6 ABOUT HERE

In the best fitting model, participant group and identifiability were the fixed effects, and the random part controlled for the variance of the different infants, listeners and stimuli nested in target words. The intercept of the model was estimated at -3.196 logits or 4% ($p < 0.001$), meaning that words are 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 single point (i.e. rating 50) on the VAS as opposed to all other points. This finding indicates that the cursor was very likely to be moved when making a judgement on the stress pattern of words. Utterances of NH infants are significantly less likely to be rated at the midpoint than those of children with CI (2%; $p < 0.001$), meaning the utterances of NH infants are more likely to be perceived as stressed. The identifiability of a stimulus slightly increases the probability of a rating at the midpoint (4%; $p < 0.001$). Adding cumulative vocabulary as a fixed effect did not improve the fit of the model. In sum, the words of CI infants are more frequently perceived to have equal stress than the words of children with NH.

5 Discussion

The aim of this study was to investigate how listeners perceive lexical stress in the early words of Dutch congenitally hearing-impaired infants with a CI and their NH peers. Acoustic research has shown a discrepancy between the prosodic characteristics of NH and CI infants (Pettinato et al., 2017). The present study investigates whether such acoustic differences are perceptually salient. This was examined by means of a listening experiment in which the disyllabic first words of NH and CI infants were presented to uninformed listeners. By moving a slider along a Visual Analogue Scale, listeners assessed two aspects of prosody: the perceived stress pattern and the degree of word stress. The first research question was whether prosody was in agreement with that of the target word. The second research question was whether listeners perceive the same degree of word stress in the first words of CI and NH children. The answers to the two research questions can be summarised as follows: (1) The words of CI and NH infants were rated in agreement with the target stress pattern as often as those of NH infants, and trochaic words were rated more frequently as such than iambic words. (2) Listeners more frequently perceived equal stress on the two syllables of the first words of infants with CI, since they are more frequently rated at the midpoint of the VAS. However, for the words that are perceived to be stressed, the degree of word stress is comparable in the two groups, and both infant groups are perceived to produce more contrast between stressed and unstressed syllables in trochees than in iambs.

Although listeners seem to perceive fewer stressed words in CI infants, they mostly indicate similar prosodic characteristics in CI and NH words. For both groups, they perceive the predominant trochaic stress pattern in the majority of the words, trochees are perceived to have a higher degree of word stress than iambs and, for both infant groups, trochaic target words are rated most frequently as such by listeners. The perceived similarities between CI and NH infants in the present study are more optimistic than would have been expected on the basis of the results from the acoustic study of Pettinato et al. (2017). That study showed a discrepancy between the CI and NH infants regarding the differentiation of intensity and F0. The perceptual results from the present investigation only partially confirm the previously attested acoustic differences: listeners who take into account all prosodic cues contributing to prominence do not hear a weaker prominence in CI words. However, the acoustic differences between the two groups materialises perceptually by scoring more towards the midpoint of the VAS. Listeners thus perceive fewer stresses in the CI group, but in general they perceive about as much differentiation between the syllables for both groups. The gradual acoustic differences between the two speaker groups thus seem to be too small to be perceived by adult listeners. This is consistent with Hide's (2013) study on 6-to-9 year old CI users: listeners considered nonsense word repetitions mostly as correctly stressed, but the acoustic measurements showed less differentiation in pitch excursion size and pitch excursion duration in CI utterances. On the one hand, this means that acoustic measurements are more likely to reveal subtle prosodic differences between two groups than a perceptual experiment. On the other hand, if the prosodic differences between groups are not large enough to be perceived, the difference may have no impact on the intelligibility of CI speech, and may thus not be as relevant as a clinical indicator. This perceptual study thus puts the acoustic differences into perspective.

It is possible that early implanted infants are able to perceive and/or process more prosodic information than could be expected given the degraded transmission of prosodic information by the implant, and that this leads to better prosody production than would be expected. Recent studies have shown that congenitally hearing-impaired infants with CI are able to perceive the lexical stress pattern of their ambient language, despite the limited transition of spectro-temporal information. Vavatzanidis et al. (2016) investigated stress perception in 17 infants with CI (9 to 50 months old, implanted at a mean age of 22 months) acquiring English during the first 6 months of implant use. In a longitudinal ERP study, they found that both CI and NH infants develop a mismatch negativity response for the iambic pattern, but not for the dominant trochaic pattern, indicating that CI children are able to discriminate the predominant stress pattern within six months after implant activation. In another recent study, Segal et al. (2016) conducted a visual habituation task to test the perception of ambient stress in 20 profoundly hearing-impaired infants with CI (12 to 33 months old, implanted under the age of 2.5 years). They found that infants with CI are able to discriminate between lexical stress patterns, but not to the same extent as their NH peers. Although CI infants have restricted auditory experience and the perception of the speech signal is degraded, they seem to be able to discriminate between stressed and unstressed syllables and perceive the predominant stress pattern of their language. The early implanted CI infants thus seem to follow similar developmental milestones as their NH peers regarding the perception of word level prominence. Since prominence perception and production are related, this might explain why listeners perceive typical production of the predominant stress pattern in the first words of CI infants.

The present results are more positive than those of studies on how listeners perceive prosody in school-aged CI children (Carter et al., 2002; Hide, 2013; Lenden & Flipsen, 2007). The study of Hide (2013) showed that listeners indicated a lower percentage of correctly imitated stress patterns in CI non-words than in NH non-words, and in the study of Carter et al. (2002), listeners only found 61% of the non-words to be repeated with the correct stress pattern. Carter et al. (2002), however, had no NH control group to compare this result with. In the present study, 63% of CI trochees and 44% of the CI iambs were considered to be accurate realizations of the target stress pattern, and there was no significant difference with the NH group (68% for trochees, 41% for iambs). However, it needs to be pointed out that there are several methodological differences between the present study and those of Hide (2013) and Carter et al. (2002). The most important is that the latter focused on prosody in non-words, whereas the present study uses spontaneous first words. Non-word repetition tasks tend to emphasize phonological differences between CI and NH speech (Nittrouer, Caldwell-Tarr, Sansom, Twersky, & Lowenstein, 2014). This makes a direct comparison between our study and those of Carter et al. (2002) and Hide (2013) difficult.

Although the results from the present study are considered to be more optimistic than the studies on prosody production by school-aged CI users, the present study also shows that CI group is perceived to have fewer words with stressed syllables. It might be the case that this is already an indicator of deprived prosody production at a later age. If this is the case, this early detection by uninformed listeners could necessitate better adapted clinical interventions to improve prosody production at a later age. This hypothesis can be investigated by more systematic longitudinal research that maps listeners' perception of prosody in CI children from their first words until school age.

The present study also sheds light on how listeners perceive prominence development in typically developing children. Listeners predominantly perceive their first words to be trochaic. Moreover, the trochaic target words are more often rated at the according side of the VAS than iambic targets. This suggests that the predominant trochaic pattern is also predominantly produced by infants who only recently started to produce lexical utterances. The fact that first words are characterised by a trochaic pattern is in line with previous studies that have argued that the production of the predominant stress pattern only becomes well established from word-use onwards (De Clerck et al., 2017; R. DePaolis, M. Vihman, & S. Kunnari, 2008). The present study also shows that vocabulary does not have an impact on how listeners perceive the degree of stress. This means that listeners do not perceive more differentiated stress when infants acquire more words. This is in line with previous acoustic studies that showed that F0, intensity and duration differentiation in disyllabic words did not increase when the vocabulary of the infants expanded (De Clerck et al., 2017; Pettinato et al., 2017). The present study did find a small vocabulary effect on the production of the stress patterns: as the infant's vocabulary increases, listeners perceive stress patterns to be in better agreement with the target stress pattern (i.e. trochaic targets produced as trochees and iambic targets produced as iambs). An increase in vocabulary size has a slight effect on how listeners perceive the production of the target stress pattern, but it does not necessarily impact listeners' perception of the degree of word stress.

An important aspect of the present study is that it focuses on word stress in spontaneous speech, recorded in a naturalistic setting and environment. This is motivated by the idea that analysis of spontaneous speech of CI users gives the most representative image of the day-to-day language production of CI infants. In the perceptual experiment, the words were presented unmanipulated in order to preserve the naturalistic nature of the dataset. This means that the segmental content of the utterances was preserved, which implies that the target words of some utterances were identifiable, whereas others were not. However, it should be pointed out that the majority of the utterances were difficult to understand, since there was no conversational context to the presented words. Since the identifiability of these words may have influenced listeners' rating behaviour, identifiability was entered as a fixed effect in the analysis. In some of the analyses, the identifiability of words did improve the fit of the model, but surprisingly it did not suggest more differentiated stress on trochaic words, which would be expected if identifiable words were considered to be more mature and thus produced with the predominant trochaic stress. Instead, a higher identifiability led to more ratings at the midpoint and a lower degree of stress. This surprising side effect can be explained by the fact that the more identifiable words were the words that are most frequently produced by the infants, such as 'mama' and 'papa'. Given the reduplicated segmental content of these words, it might be the case that the prominence in these words is also less differentiated. More importantly, the finding that the trochaic pattern is predominant in the ratings does not seem to be attributed to the identifiability of the words. An alternative is to low pass filter the stimuli. By erasing the segmental content and only preserving the suprasegmental features, the words would be unidentifiable. However, since identifiability did not impact listeners' perception of the stress pattern, low pass filtering the stimuli was unlikely to influence the results. In sum, this study shows that reliable results are obtained when spontaneous utterances are presented to uninformed listeners on a VAS.

Conclusion

From this study, it can be concluded that listeners perceive typical word stress production in the first words of congenitally hearing-impaired infants who received a CI before 20 months. Although listeners perceive fewer clear stresses in the first words of CI infants, the assessment of prosody in CI and NH words was very similar. Listeners do not indicate less gradual modulation in the words of CI infants, and the first words of CI infants are as often rated in agreement with the stress pattern of the attempted target word. In conclusion, this study shows that acoustic differences between CI and NH infants' stress production are not necessarily perceptually salient, and it suggests that ratings from uninformed listeners can be very useful as a clinical indicator to evaluate CI speech.

Acknowledgements:

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. Our special thanks go to the families and infants that participated in the study and to K. Schauwers, I. Molemans, R. Van den Berg and L. Van Severen for collecting the CLiPS Child Language Corpus.

References

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

Bates, D , Maechler, M , Bolker, BM , & 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

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

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

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. doi: 10.1017/s0305000916000532

DePaolis, R., Vihman, M., & Kunnari, S. (2008). Prosody in production at the onset of word use: A cross-linguistic study. *Journal of Phonetics, 36*, 406-422.

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 and Neurotology, 37*(2), e82-e95. doi: 10.1097/mao.0000000000000915

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. doi: 10.1682/jprd.2007.08.0118

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 and Speech, 1*(2), 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. doi: 10.1121/1.1785611

'tHart, J. T., Collier, R., & Cohen, A. (1990). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.

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.

Hunter, C. R., Kronenberger, W. G., Castellanos, I., & Pisoni, D. B. (2017). Early Postimplant Speech Perception and Language Skills Predict Long-Term Language and Neurocognitive Outcomes Following Pediatric Cochlear Implantation. *Journal of Speech, Language, and Hearing Research, 60*(8), 2321-2336. doi: 10.1044/2017_jslhr-h-16-0152

Julien, H., & Munson, B. (2012). Modifying Speech to Children Based on Their Perceived Phonetic Accuracy. *Journal of Speech, Language, and 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 and Hearing Disorders, 52*(1), 64-75.

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

Levine, Dani, Strother-Garcia, Kristina, Golinkoff, Roberta Michnick, & Hirsh-Pasek, Kathy. (2016). Language Development in the First Year of Life: What Deaf Children Might Be Missing

Before Cochlear Implantation. *Otology and Neurotology, 37*(2), e56-e62. doi: 10.1097/mao.0000000000000908
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. doi: 10.1121/1.1908095
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. doi: 10.1016/j.jcomdis.2016.07.001
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.
Mertens, P. (2001). Fonilex. Retrieved 14 august 2014 http://bach.arts.kuleuven.be/fonilex/
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.
Moore, B. C. J. (2003). Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants. *Otology and Neurotology, 24*(2), 243-254. doi: 10.1097/00129492-200303000-00019
Most, T., & Peled, M.s. (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), 350-361.
Munson, B., Schellinger, S. K., & Carlson, K. U. (2012). Measuring speech-sound learning using visual analog scaling. *SIG 1 Perspectives on Language Learning and Education, 19*, 19-30. doi: doi:10.1044/lle19.1.19
Nittrouer, S., Caldwell-Tarr, A., Sansom, E., Twersky, J., & Lowenstein, J. H. (2014). Nonword repetition in children with cochlear implants: a potential clinical marker of poor language acquisition. *Am J Speech Lang Pathol, 23*(4), 679-695. doi: 10.1044/2014_ajslp-14-0040
O'Halpin, R. (2010). *The perception and production of stress and intonation by children with cochlear implants.* (Doctoral dissertation), University College London, London.
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 and Hearing, 38*(4), 475-486. doi: 10.1097/aud.0000000000000406
Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crepin, G. (1988). Fetal hearing. *European Journal of Obstetrics and Reproductive Biology, 28*, 191-212.
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
Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology, 33*(1), 3-11.
Schauwers, K. (2006). *Early speech and language development in deaf children with a cochlear implant: A longitudinal investigation.* (Doctoral dissertation), University of Antwerp, Antwerp.
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.
Segal, O., Houston, D., & Kishon-Rabin, L. (2016). 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
Titterington, J., Henry, A., Kramer, M., Toner, J. G., & Stevenson, M. (2006). An investigation of weak syllable processing in deaf children with cochlear implants. *Clinical Linguistics and Phonetics, 20*(4), 249-269. doi: 10.1080/02699200400015291
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, and Hearing Research, 48, 853-867.
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.
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.
Vavatzanidis, N. K., Mürbe, D., Friederici, A. D., & Hahne, A. (2016). The Perception of Stress Pattern in Young Cochlear Implanted Children: An EEG Study. *Frontiers in Neuroscience, 10*, 68. doi: 10.3389/fnins.2016.00068
Verhoeven, J. (2005). Belgian Standard Dutch. *Journal of the International Phonetic Association, 35*, 243-247.
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.

Table 1: Auditory characteristics of the CI children in the corpus.

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

Legend: PTA = Pure Tone Average at the age of 2; dBHL = decibel hearing level; HA = Hearing Aid; CI = Cochlear Implant

Table 2: Recording information on the individual children.

ID	Gender	Age start (y;mm.dd)	Age end (y;mm.dd)	# utterances
CI-1	F	2;02.26	2;09.27	93
CI-2	F	1;10.19	2;04.27	43
CI-3	F	1;11.22	2;06.09	163
CI-4	M	1;11.23	3;06.16	24
CI-5	M	1;08.19	2;05.27	28
CI-6	M	1;06.00	2;04.26	55
CI-7	F	1;07.10	2;04.16	59
CI-8	F	2;02.08	2;10.13	35
CI-9	F	1;05.21	1;09.20	29
CI TOTAL	67% F	Mean: 1;09.17	Mean: 2;06.10	Total: 529 (mean: 28.78; SD: 44.59)
NH-1	M	1;03.10	2;00.01	115
NH-2	M	1;02.05	2;00.02	49
NH-3	M	1;03.04	1;11.05	97
NH-4	M	1;01.28	1;11.30	104
NH-5	F	1;01.28	1;11.29	46
NH-6	M	1;07.01	2;00.04	4
NH-7	F	1;06.04	1;11.27	48
NH-8	F	1;04.04	2;00.04	30
NH-9	F	1;03.28	1;09.28	67
NH TOTAL	44% F	Mean: 1;03.10	Mean: 1;11.06	Total: 560 (mean: 62.22; SD: 36.80)
OVERALL TOTAL	56% F	Mean 1:06.13	Mean: 2;02.24	Total: 1089 (mean: 60.50; SD:40.70)

Legend: CI = Cochlear Implanted infants; NH = Normally hearing infants; SD = Standard Deviation

Table 3: Descriptive statistics of the perceived stress pattern.

	Proportion of the target stress pattern per group		Proportion of ratings in agreement with the target stress pattern	
	NH	CI	NH	CI
Trochee	0.86	0.89	0.68	0.63
Iamb	0.14	0.11	0.41	0.44

Legend: CI = Cochlear Implanted infants; NH = Normally hearing infants

Table 4: Generalised mixed model with agreement between ratings dependent variable.

Random Effects	**Variance**	**Standard deviation**			
Target words	0.059	0.243			
Participant identity	0.003	0.585			
Residual	0.042	0.205			
Fit of the model (AIC)	-8335.4				
Fixed Effects	**Estimate (ß)**	**Standard Error**	**df**	**t**	**p**
Intercept	0.238	0.027	62	8.732	< 0.001
Participant group (NH)	-0.027	0.029	20	-0.926	0.366
Stress pattern of target word (trochee)	0.413	0.008	2679	52.970	< 0.001
Identifiability	0.023	0.001	27874	19.172	< 0.001
Cumulative vocabulary	0.0001	0.00004	27227	2.152	0.030
Participant group (NH) x cumulative vocabulary	0.0004	0.00005	27389	7.903	< 0.001

Legend: Participant group = NH or CI.

Table 5: Generalised mixed model with degree of stress as dependent variable.

Random Effects	**Variance**	**Standard deviation**			
Stimuli nested in target words	11.115	3.334			
Rater identity	29.452	5.427			
Participant identity	0.396	0.630			
Residual	44.048	6.637			
Fit of the model (AIC)	195757.9				
Fixed Effects	**Estimate (ß)**	**Standard Error**	**df**	**t**	**p**
Intercept	14.352	0.647	111	22.159	< 0.001
Participant group (NH)	0.651	0.387	14	1.682	0.111
Identifiability	-0.187	0.073	1054	-2.540	0.011
Rating location (trochaic side)	1.789	0.109	27432	16.363	< 0.001

Legend: Participant group = NH or CI.

Table 6: Logistic regression analysis in the form of a generalised mixed model with the likelihood to be rated at the midpoint of the VAS as dependent variable.

Random Effects	**Variance**	**Standard deviation**		
Stimuli nested in target words	1.587	1.260		
Rater identity	1.307	1.143		
Participant identity	0.046	0.190		
Fit of the model (AIC)	15241.5			
Fixed Effects	**Estimate (ß)**	**Standard Error**	**t**	**p**
Intercept	-3.196	0.163	-19.630	< 0.001
Participant group (NH)	-0.609	0.135	-4.503	< 0.001
Identifiability	0.103	0.031	3.323	< 0.001

Legend: Participant group = NH or CI.

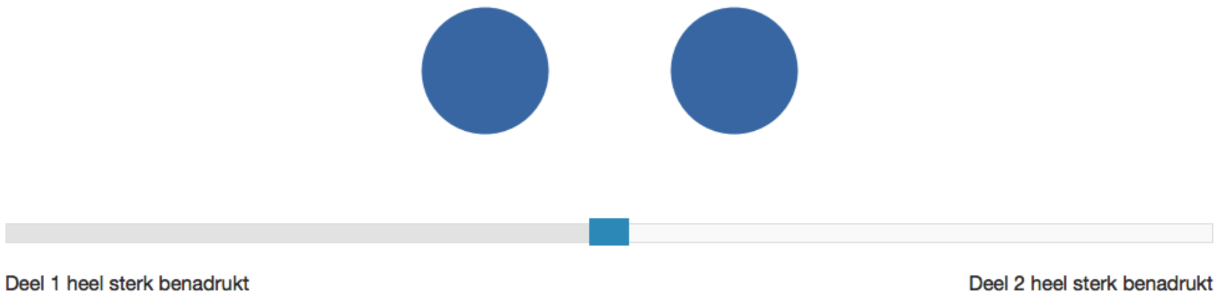


Figure 1a: equally stressed syllables

Figure 1b: stress on the first syllable

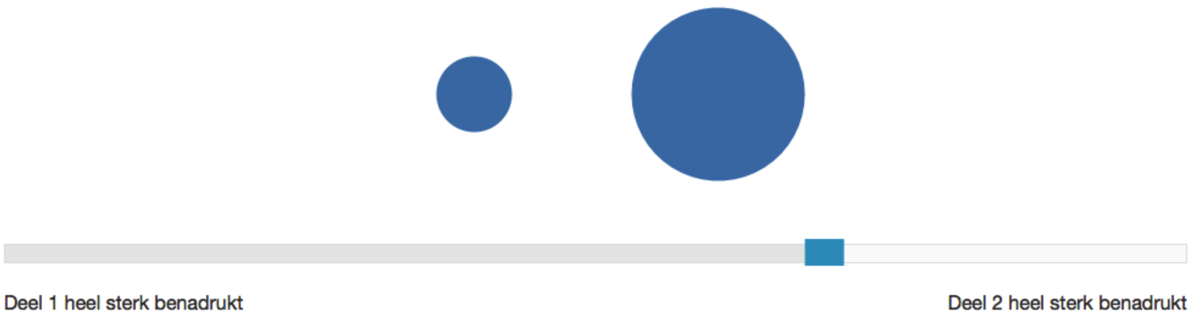


Figure 1c: stress on the second syllable

Figure 1: Print screen of three different rating positions on the sliding bar. Legend: "Deel 1/2 heel sterk benadrukt" = "A very prominent part 1/2"

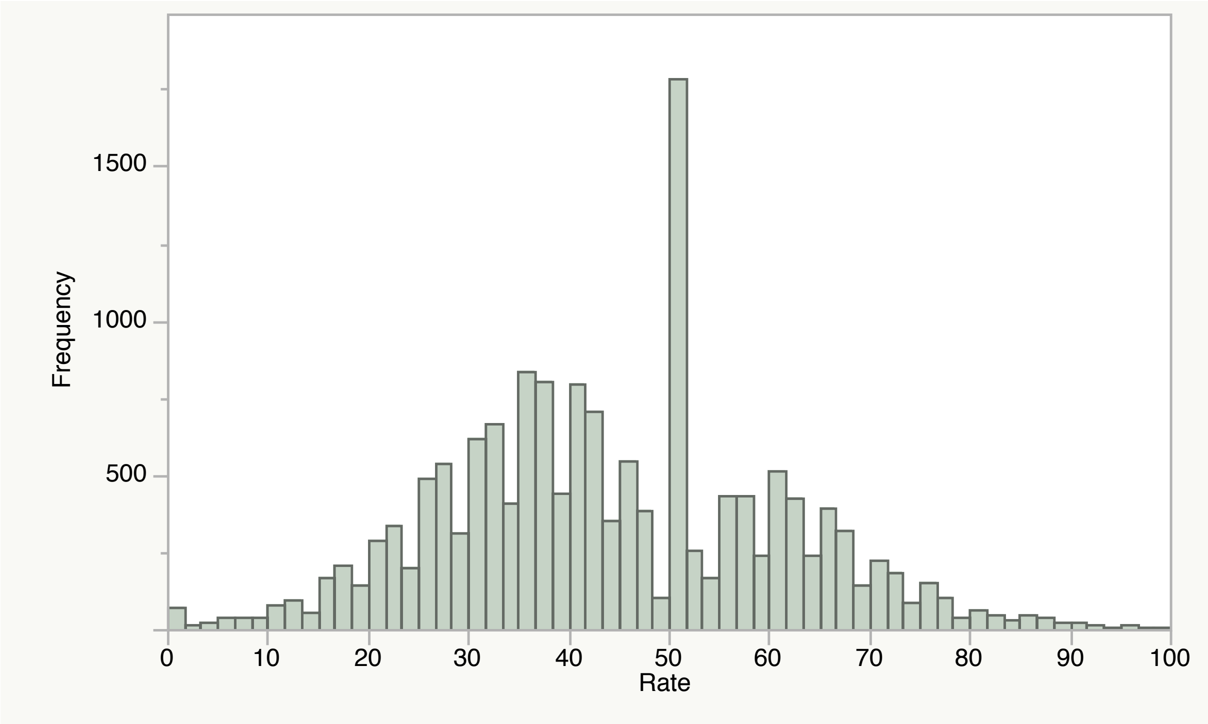


Figure 2a: Distribution of all ratings in CI group

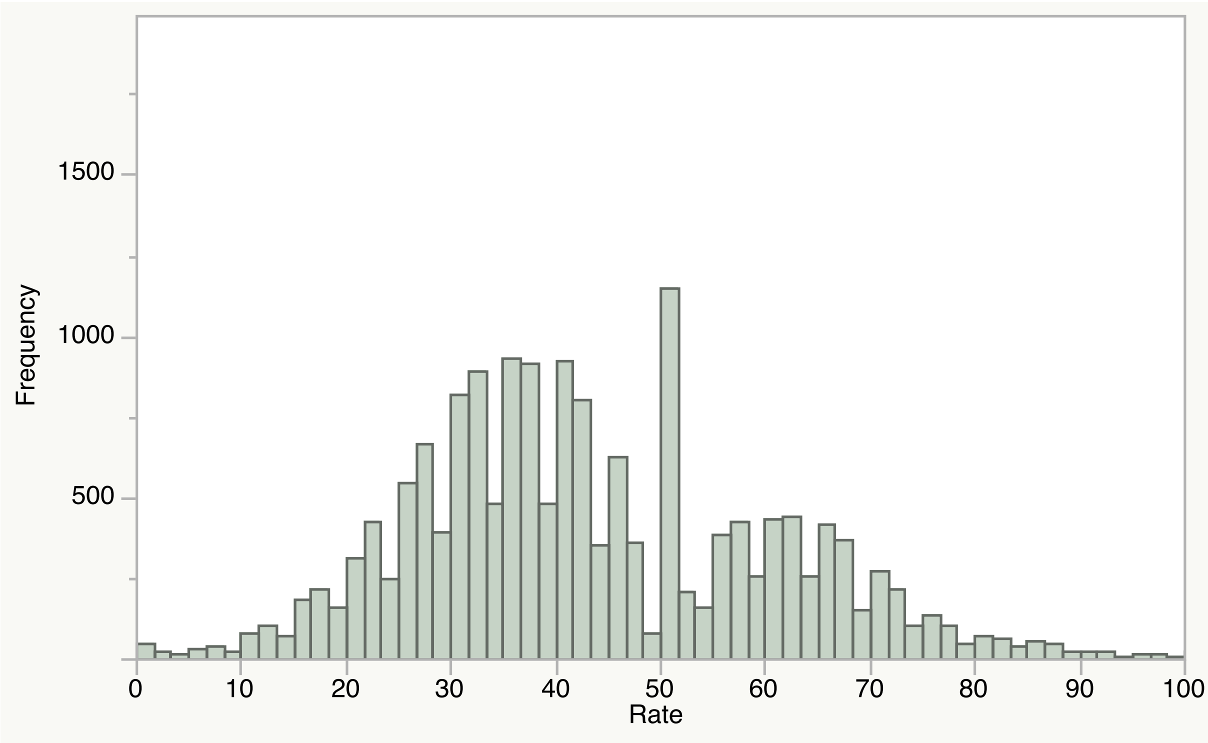


Figure 2b: Distribution of all ratings in NH group

Figure 2: Distribution of all ratings per participant group.

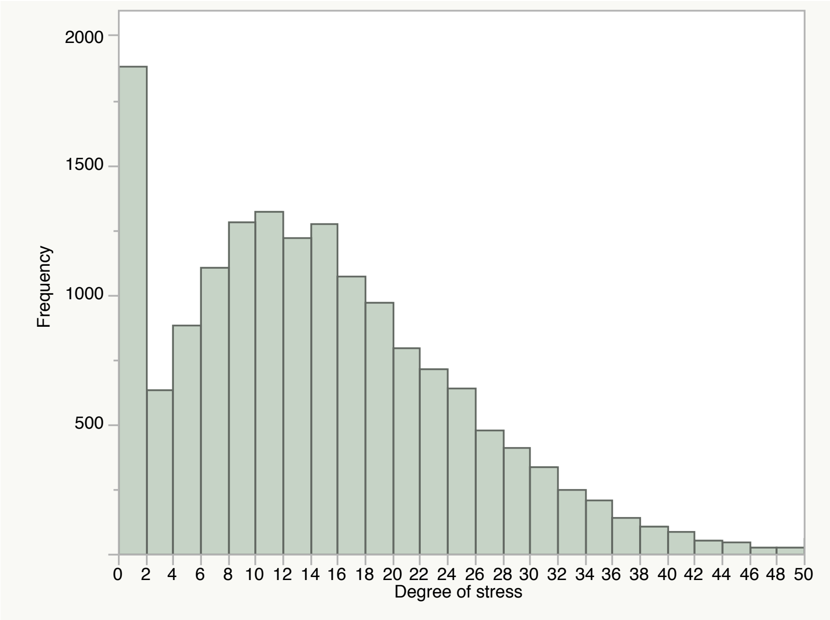


Figure 3a: Distribution of the degree of stress in CI group

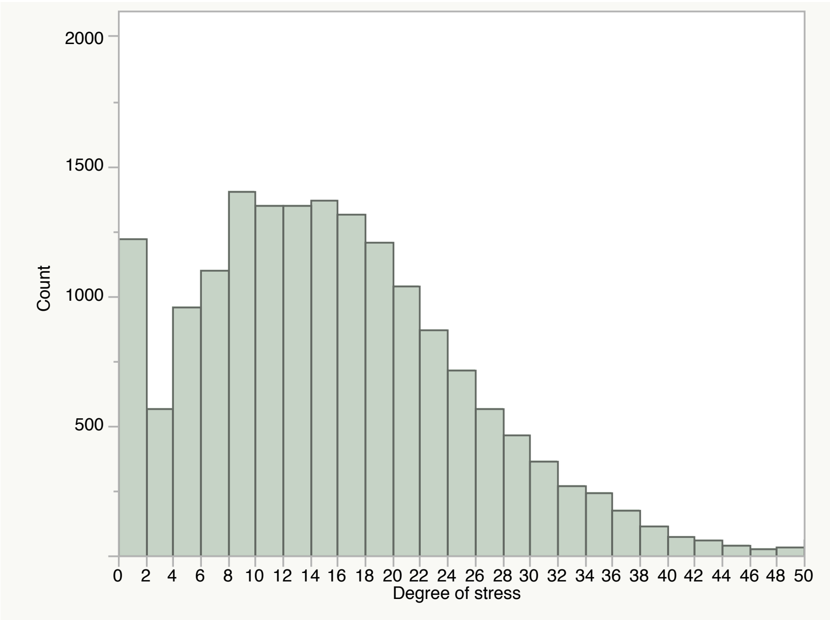


Figure 3b: Distribution of the degree of stress in NH group

Figure 3: Distribution of the degree of stress per participant group. Legend: 0 = midpoint of VAS; 50 = extremes of VAS

Supplemental material 1: *recruitment questionnaire*

- Age?
- Sex?
- Do you have hearing problems? (Tinitus, hearing loss,…)? Remarks?
- Does somebody in your family have hearing problems? Remarks?
- Did you go for swimming the past 24 hours? Remarks?
- Did you go to a concert the past 24 hours? Remarks?
- Do you have a cold or an ear infection? Remarks?
- Did you recently had a cold or an ear infection? Remarks?
- Do you have other problems that might lead to hearing loss? Remarks?
- Have you ever been diagnosed with a developmental disorder (such as: autism, dyslexia, dyspraxia, specific language impairement,…)? Remarks?
- Do you have a linguistic background? Remarks?
- Do you have a phonetic background? Remarks
- How often do you come in to contact with infants between zero to two years old?
    a. Never
    b. Once a year
    c. Several times a year
    d. Onse a month
    e. Several times a month
    f. Every week
    g. Daily

Supplemental material 2: *Overview of the types and tokens of target words produced per participant group. Legend:* ***Bold words are included in the subset with only disyllabic target words***

Target word (translation)	CI	NH
Aaitje (diminutive of 'stroke')	**2**	**1**
Aandoen ('put on')	**0**	**1**
Aap ('monkey')	2	0
Apen ('monkeys')	**0**	**1**
Aardbei ('strawberry')	**0**	**1**
Allemaal ('all')	1	2
Alsjeblieft ('please')	0	1
Amber (a name)	**7**	**0**
An-Sofie (a name)	1	0
Ander ('other')	**0**	**2**
Anton (a name)	**6**	**0**
Appel ('apple')	**13**	**16**
Appelsien ('orange')	1	1
Auto ('car')	**7**	**35**
Auto's ('cars')	**0**	**1**
Baby ('baby')	**1**	**12**
Badje ('little bath')	**0**	**3**
Bal ('ball')	0	1
Ballon ('balloon')	**4**	**5**
Baltazar (a name)	0	1
Banaan ('banana')	**2**	**3**
Banaantje (diminutive of 'banana')	0	2
Bang ('afraid')	1	1
Beertje (diminutive of 'bear')	**1**	**7**
Beestje (diminutive of 'animal')	**0**	**2**
Beker ('cup')	**0**	**2**
Betalen ('to pay')	0	1
Beter ('better')	**0**	**1**
Binnen ('inside')	**0**	**1**
Bloemen ('flowers')	**0**	**3**
Blokje (diminutive of 'block)	**0**	**1**
Blokken ('blocks')	**1**	**1**
Blote ('naked')	**0**	**1**
Boekje ('booklet')	**0**	**5**
Boempatat (onomatopea)	0	1
Boke (shortened form of botherham, 'sandwich')	**0**	**4**
Boma (a name)	**1**	**0**
Bompa ('grandpa')	**1**	**0**
Bootje (diminutive of 'boat')	**3**	**3**
Boterham ('sandwich')	0	1
Boven ('upstairs')	**0**	**2**
Bran (a name)	0	1
Brandweer ('firefighter')	**1**	**0**
Bravo ('well done')	**1**	**0**

Brilletjes (diminutive of 'glases')	0	1
Broodje (diminutive of 'sandwich')	**0**	**1**
Bubu (a name)	**8**	**0**
Buggy ('buggy')	**0**	**1**
Buik ('belly')	1	0
Buiten ('outside')	**2**	**1**
Bukken ('to bend over')	**1**	**0**
Bumba (a name)	**0**	**17**
Cadeautje (diminiutive of 'present')	0	1
Camera ('camera')	0	1
Cavia ('guinea pig')	**0**	**1**
Choco ('chocolate paste')	**0**	**1**
Chocolaatje (diminutive of 'chocolate')	0	1
Coca ('Coca Cola')	**1**	**0**
Cola ('Coca Cola')	**1**	**0**
Computer ('computer)	0	1
Daar ('overthere')	0	11
Dada ('bye bye')	**5**	**9**
Danku ('thank you')	**4**	**0**
Dansen ('dance')	**0**	**3**
Dat ('that')	0	1
David (a name)	**5**	**0**
Deksel ('lit')	**0**	**1**
Deur ('door')	0	1
Deurtje (diminutive of 'door')	**0**	**1**
Deze ('this')	**0**	**6**
Die ('that')	0	1
Dikke ('fat')	**0**	**1**
Dodo ('nappy')	**5**	**1**
Dokter ('doctor')	**1**	**0**
Donker ('dark')	**0**	**1**
Dood ('dead')	1	0
Drinken ('to drink')	**0**	**2**
Diseme (a name)	**2**	**0**
Eekhoorn ('squirrel')	**0**	**1**
Eenden ('ducks')	**1**	**0**
Eendje (diminutive of 'duck')	**0**	**9**
Eendjes (diminutive of 'ducks')	**0**	**1**
Egan (a name)	**0**	**1**
Egel ('hedgehog')	**3**	**0**
Eikels ('acorns')	**0**	**1**
Eitje (diminutive of 'egg')	**0**	**1**
Emma (a name)	**26**	**0**
Eten ('to eat' or 'food')	**6**	**11**
Ezel ('donkey')	**0**	**2**
Fietsen ('to cycle' or 'bikes')	**1**	**0**
Foko (a name)	**0**	**1**

Foto ('picture')	**0**	**1**
Gedaan ('finished')	**0**	**6**
Geitje (diminutive of 'goat')	**0**	**1**
Geven ('to give')	**0**	**1**
Gezet ('sat')	**0**	**1**
Glijbaan ('slide')	**2**	**5**
Graafmachine ('digging machine')	0	1
Hallo ('hello')	**4**	**0**
Handje (diminutive of 'hand')	**0**	**1**
Handjes (diminutive of 'hands')	**3**	**0**
Hebben ('to have')	**1**	**0**
Helicopter ('helicopter)	**0**	**4**
Helpen ('ho help')	**0**	**3**
Hertjes (diminutive of 'reindeer')	**1**	**0**
Hetzelfde ('the same')	**3**	**0**
Hier ('here')	**1**	**0**
Hond ('dog')	**2**	**0**
Hondje ('doggy')	**3**	**4**
Hopla (name of baby tv-show character)	**2**	**4**
Huisje (diminutive of 'house')	**0**	**2**
Ijsje (diminutive of 'icecream')	**0**	**5**
Ik ('I')	**4**	**2**
Inge ('a name')	**0**	**1**
Is dat ('what's that')	**1**	**1**
Italië ('Italy')	**1**	**0**
Ja ('yes')	**1**	**0**
Kabauter Plop (name of baby tv-show character)	**0**	**1**
Kaka ('poopoo')	**7**	**7**
Kalkoen ('turkey')	**0**	**3**
kan niet ('can't')	**0**	**1**
Kapot ('broken')	**14**	**2**
Kasper (a name)	**5**	**0**
Keuken ('kitchen')	**1**	**0**
Kijk ('look')	**0**	**1**
Kijken ('to look')	**26**	**6**
Kikker ('frog')	**0**	**7**
Kindje (diminutive of 'child')	**7**	**10**
Kippen ('chickens')	**1**	**1**
Kiwi ('kiwi')	**0**	**1**
Kleertje (diminutive of 'cloth')	**0**	**2**
Clement (a name)	**0**	**1**
Knippen ('to cut')	**1**	**0**
Koala ('koala)	**1**	**0**
Koek ('cookie')	**1**	**0**
Koeken ('cookies')	**0**	**1**
Koekje ('cookie')	**0**	**1**
Koffie ('coffee')	**3**	**0**

Konijn ('rabbit')	**5**	**5**
Konijntje ('bunny')	**0**	**1**
Koud ('cold')	**1**	**0**
Kousje (diminutive of 'sock')	**1**	**0**
Krikrak (onomatopea)	**1**	**0**
Kristof (a name)	**10**	**0**
Krokodil ('crocodile')	**0**	**2**
Kruipen ('to crawl')	**0**	**1**
Kuiken ('chick')	**0**	**2**
Kunnen ('can')	**0**	**1**
Lala (onomatopea)	**0**	**2**
Lange ('long')	**0**	**1**
Lara (a name)	**0**	**4**
Larsje (a name)	**0**	**1**
Leander (a name)	**1**	**0**
Lela (a name)	**0**	**1**
Lepel ('spoon)	**1**	**0**
Leven ('to live')	**0**	**1**
Lieve (a name)	**0**	**1**
Limonade ('lemonade')	**0**	**2**
Lopen ('to run')	**4**	**0**
Maken ('to make')	**6**	**1**
Mama ('mommy')	**38**	**21**
Man ('man')	**1**	**0**
Marleen (a name)	**1**	**0**
Meisje ('girl')	**0**	**1**
Melk ('milk')	**1**	**0**
Melkje (diminutive of 'milk')	**0**	**1**
Meme ('granny')	**0**	**1**
Meneer ('mister')	**0**	**1**
Mieke (a name)	**3**	**0**
Moeke ('mommy')	**0**	**2**
Mondje (diminutive of 'mouth')	**0**	**1**
Morgen ('tomorrow')	**0**	**1**
Muisje (diminutive of 'mouse')	**1**	**1**
Neen ('no')	**1**	**0**
Negen ('nine')	**1**	**0**
Neushoorn ('rhino')	**1**	**1**
Nick ('a name')	**0**	**1**
Nijlpaard ('hippopotamus')	**0**	**1**
Nina (a name)	**0**	**1**
Nono (a name)	**1**	**0**
Olifant ('elephant')	**0**	**3**
Oma ('granny')	**12**	**1**
Omdraaien ('to turn around')	**1**	**0**
Onder ('under')	**1**	**0**
Onki (a name)	**0**	**1**
Oogjes (diminutive of 'eyes')	**0**	**1**
Ooievaar ('stork')	**1**	**0**

Oortjes (diminutive of 'ears')	**0**	**1**
Opa ('grandpa')	**7**	**0**
Open ('open')	**17**	**7**
Opruimen ('to clean up')	**1**	**0**
Paardje (diminutive of 'horse')	**8**	**8**
Pablo (a name)	**0**	**1**
Pakje ('gift')	**1**	**0**
Pakjes ('gifts')	**1**	**0**
Pakken ('pick up')	**1**	**4**
Panda ('panda')	**1**	**0**
Pannen ('pans')	**0**	**1**
Papa ('daddy')	**57**	**35**
Papegaai ('parrot')	**2**	**4**
Papiertje (diminutive of 'paper')	**1**	**0**
Paraplu ('umbrella')	**1**	**0**
Patat ('potato')	**0**	**1**
Paulientje (a name)	**1**	**0**
Peter (a name)	**1**	**0**
Piano ('piano')	**0**	**3**
Pinguin ('penguin')	**0**	**2**
Pipi ('pee')	**2**	**2**
Pipo (a name of a tv clown)	**10**	**0**
Poes ('cat')	**3**	**2**
Poesje ('kitty')	**1**	**10**
Poetsen ('to clean')	**0**	**2**
Pop ('doll')	**1**	**0**
Potje ('potty')	**0**	**1**
Proper ('clean')	**1**	**0**
Puzzel ('puzzle')	**0**	**1**
Raampje (diminutive of 'window')	**0**	**3**
Rechtstaan ('to stand up')	**0**	**2**
Renate (a name)	**0**	**2**
Rijden ('to drive')	**0**	**2**
Roxanne (a name)	**1**	**0**
Sant (a name)	**2**	**0**
Schaap ('sheep')	**0**	**1**
Scheetje (diminutive of 'fart')	**0**	**1**
Schoenen ('shoes')	**2**	**0**
Schuifaf ('slide')	**0**	**3**
Severin (a name)	**1**	**0**
Sien (a name)	**2**	**0**
Slapen ('to sleep')	**3**	**4**
Sokken ('socks')	**0**	**1**
Speeltuin ('playground')	**1**	**1**
Spelen ('to play')	**1**	**1**
Stappen ('to walk')	**0**	**2**
Steven (a name)	**1**	**0**
Stoeltje (diminutive of 'chair')	**1**	**0**
t-shirt ('t-shirt')	**1**	**0**

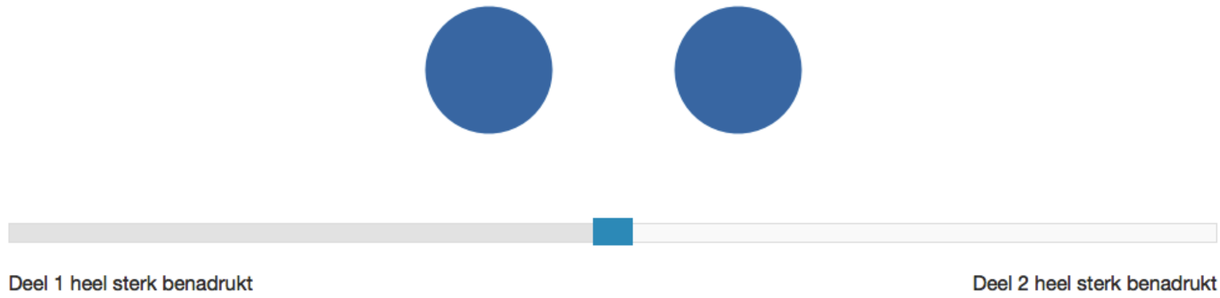
Taartje (diminutive of 'cake')	**0**	**2**
Tafel ('table')	**2**	**6**
Tandjes (diminutive of 'teeth')	**0**	**1**
Tanekin (a name)	**2**	**0**
Tekenen ('to draw')	**1**	**0**
Telefoon ('telephone')	**1**	**1**
Tessa (a name)	**3**	**0**
Tiktak (onomatopea for a ticking clock, name of a tv show)	**1**	**1**
Toedoen ('to close')	**2**	**1**
Tomaat ('tomato')	**1**	**0**
Traktor ('tractor')	**0**	**3**
Trommel ('drum')	**0**	**1**
Trompet ('trumpet')	**0**	**1**
Tuinman ('gardener')	**1**	**0**
Tutje (diminutive for 'pacifier')	**2**	**1**
Tv ('tv')	**0**	**2**
Varken ('pig')	**5**	**3**
Vast ('firm' or 'closed')	**0**	**1**
Vera (a name)	**1**	**0**
Visje (diminutive of 'fish')	**0**	**1**
Vlieger ('airplane')	**0**	**1**
Vliegtuig ('airplane')	**2**	**9**
Vlinder ('butterfly')	**2**	**1**
Voetjes (diminutive of 'feet')	**0**	**1**
Vogel ('bird')	**5**	**3**
Wafwaf (onomatopea for a dog's barking)	**0**	**1**
Wakker ('awake')	**0**	**1**
Wassen ('to wash')	**1**	**0**
Wat is dat ('what's that')	**19**	**0**
Water ('water)	**1**	**6**
Werken ('to work')	**1**	**0**
Wie is dat ('who's that')	**3**	**0**
Winnie the Pooh ('Winnie the Pooh')	**2**	**0**
Wolkje (diminutive of 'cloud')	**0**	**5**
Wortel ('carrot')	**0**	**1**
Zebra ('zebra')	**0**	**2**
Zetel ('sofa')	**0**	**1**
Zingen ('to sing')	**0**	**1**
Zitten ('to sit')	**2**	**0**
Zoeken ('to search)'	**0**	**1**
Zwemmen ('to swim')	**0**	**6**
Total number of words	**529**	**560**
No target transcription	**14**	**12**
Total number of types	**139**	**191**
Total number of tokens	**515**	**548**
Number of disyllabic types	**102**	**156**
Number of disyllabic tokens	**443**	**488**

Supplemental material 3: *Screencast of the perceptual experiment*

Supplemental material 4: *Overview of the instructions to the perceptual experiment*

In this experiment you will get to hear utterances produced by infants. These utterances consist of two parts (for instance “TI TI”). These two parts can be prominent in different gradations. For instance, the first part can be a bit, slightly or very prominent. The same can be the case for the second part. The prominence 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:



If you want to listen again to the utterance you push the button below:

(Herbeluister = ‘repeat’)

In total you can listen to the utterance three times.

Before the experiment starts, you will see and hear six clear example utterances. You will hear an utterance and see how the slider is moved to the right position. After these six examples you can have a try yourself on 20 utterances in which the prominence 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 prominent part correctly. You can try again when this message occurs. If the bar is placed on the correct side of the axis you can continue to the next utterance.

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

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

IMPORTANT!

During the trial phase the prominence will be clearly located on either the first or the second part of the utterance. This training is designed to familiarise you with the fact that two parts of the utterance carry different prominence. In the utterances presented in the actual experiment the prominence 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 utterances 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 utterances.
    - With the first 10 utterances you can only continue to the next stimulus if the bar is correctly positioned.
    - With the last 10 utterances 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.

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