# The production of word stress in babbles and early words: a comparison between normally hearing infants and infants with cochlear implants.

Michèle Pettinato[1], Ilke de Clerck[1], Jo Verhoeven[1,2], Steven Gillis[1]

[1] Antwerp University, [2] City University London
michele.pettinato@uantwerpen.be, Ilke.DeClerck@uantwerpen.be, Jo.verhoeven@city.ac.uk, steven.gillis@uantwerpen.be

## ABSTRACT

There is evidence that infants are able to manipulate cues to word stress as early as babbles. For children with cochlear implants (CI), word stress production may pose difficulties since the CI does not provide enough detail for adequate perception of f0 or intensity changes. This study is a longitudinal investigation of pitch, intensity and duration in disyllabic babbles and first words by normally hearing (NH) and children with CI. Both groups had smaller acoustic differences on babbles than on words, and children with CI made smaller differences in pitch and intensity than the NH group. A marked increase in acoustic differentiation was seen in the NH group on words, especially for pitch. Although there was a trend in the same direction in the CI group, the same shift was not observed in their lexical productions. Implications for language processing in this population and theories of language acquisition are considered.

**Keywords**: word stress, cochlear implants, babbling, first words.

## 1. INTRODUCTION

This study examined the development of word stress production in disyllabic babbles and words in the early spontaneous speech of children with cochlear implants (henceforth CI) and normally hearing (NH) children. Three acoustic cues were investigated: f0, intensity and duration of vowels.

Word stress is highly salient [18] and it is one of the earliest aspects of speech to be discriminated by infants. Word stress also guides the production of early word forms [9], and there is evidence that children can manipulate the cues to word stress from early on: Davis et. al [7] report 7-14 month old infants were able to use f0, intensity and duration to create prominence differences in disyllabic babbles, and DePaolis et al. [8] showed further convergence of cues towards adult, language-specific patterns in the disyllables of 10-18 month old infants. Nevertheless, the phonetics of children's stress patterns continues to be fine-tuned up to 7 years [1].

For children with CI, the acquisition of word stress may pose difficulties since the spectral and temporal resolution provided by the implant does not provide enough detail for adequate perception of F0 [10][21] or changes in intensity [10][19]. Durational properties of syllables seem to be available to listeners with a CI [19][21]. Whilst f0 is not available as a cue for prosody, temporal aspects of pitch may nevertheless aid perception [5]. Indeed Torppa et al. [24] found age-equivalent stress perception in children with CI who had musical training. The complex picture for perception is mirrored in production. In a study of conversational speech samples of six children with CI, Lenden & Flipsen [16] noted abnormalities in word and sentence stress, which sounded 'excessive, equal or misplaced' (p.75), whilst measures of phrasing, voice quality and pitch were relatively preserved. In contrast, in nonsense repetition tasks children with CI had better performance on word stress than segmental accuracy [22]. Finally, although the stress productions of 6-8 year old Dutch speaking children with CI were mostly rated correct, they made less distinct differences in acoustic cues between stressed and unstressed syllables [12]. The present study was therefore concerned with a younger age group, at the transition from babbling to first words. Two research questions were examined. Firstly, we were interested in whether the relative success described in [12] would also be evident at earlier stages, i.e. whether the CI group used the acoustical cues in a similar way to the NH group. In contrast to most of the studies reviewed, the children in the present study were implanted before the age of two. We therefore supposed that this group would essentially be able to render word stress, similarly to the advantage in perception described in [24]. Developmental aspects were a second focus: changes over time and in particular, the effect of the emergence of first words. Because words have a clear adult target, and the child is using them with communicative intent, it was hypothesized that stress differences may become clearer in words. Since children with CI receive little early auditory input and continue to receive degraded input from other's and their own speech, it was also hypothesized that differences between groups may become accentuated in words.

## 2. MATERIALS AND METHODS

### 2.1. Participants

The participants were 9 children using CI and 9 NH peers. The children with CI were implanted before the age of two (M=12 mths; SD= 5). Recordings for the NH group started at a M= 6 mths (SD= 0.72) until M = 22 mths (SD= 3); for the CI group they started at M= 17 mths (SD= 4), ending at M= 24 mths (SD= 4). For the control group, normal language development was verified through the Dutch version of the CDI ("N-CDI") [27]. The children were matched on developmental stages: recordings for a child were included from the onset of babbling as determined by a True Canonical Babbling Ratio of 0.15.[4] until s/he reached a cumulative vocabulary of 200 words[13].

### 2.2. Corpus

The corpus consisted of monthly 20 min recordings of spontaneous interactions between the children and their caretakers in their home. The criteria for distinguishing words (lexical items) from babbles (prelexical items) were based on [2]. Recordings were first transcribed following CHAT conventions [3]. The chat files were then converted to PRAAT [4] TextGrids using the CHAT2PRAAT function in CLAN [17].

### 2.3. Data selection

Disyllabic words and babbles were selected from the waveform using auditory and visual criteria and tagged on the TextGrids. A sound was considered to be a disyllable when it consisted of two vocalic phases minimally separated by a clear consonantal phase. The inclusion criteria for the disyllables were adapted from [5]: disyllables that were single-word utterances had to be clearly separated from surrounding speech by a silence of at least 400 ms with an intersyllabic pause of less than 400 ms. Disyllables which were clearly single items in spite of the fact that they were separated by less than 400 ms from surrounding speech or occurred within the same intonation contour or breath group [6] were tagged to signal that they were part of a multi-word utterance. For words, these could occur with other recognisable lexical material or surrounding babble. However, babbled disyllables in a multiword context could only be identified as such if they co-occurred with lexical material, since it was difficult to determine boundaries with other babbles. Disyllables were excluded if there was concurrent speech or noise or if they were produced with a creaky, breathy, excessively loud or whispered voice. The disyllables were further segmented into consonants and vowels, since the acoustic measurements were conducted in the vocalic portion of each syllable. These annotations were carried out on the basis of the segmentation guidelines provided in [7]. The final dataset consisted of 2519 disyllables, 443 (17.6%) of which came from a multiword context. There were 785 babbles (CI=253; NH=532) and 1734 words (CI=885; NH=849). For the reliability, 10% (*N*=250) of the data was re-annotated by a second researcher. The Pearson's correlation between segment durations was 0.99.

### 2.4. Acoustic analysis

Duration, intensity and f0 were measured in the tagged vocalic portions using PRAAT. Duration (in ms) was measured from the start to the end of each vowel. Intensity was measured in dB as the mean energy averaged over the total number of analysis frames in the vowel. F0 was determined by means of the PRAAT autocorrelation method and expressed in Hz as the mean F0 averaged over the total number of analysis frames in the vowel. The analysis settings were adjusted to child speech, i.e. f0 range was set at 150-800 Hz and intensity range was set at 0-100 dB. Subsequently, duration ratio (1) and intensity ratio (2) for the two vowels in each disyllable were calculated. Furthermore, the perceptual pitch distance between the first and the second vowel in semitones was calculated using the following formula (3).

(1) $\text{Duration ratio} = \text{Duration}_{V1}/\text{Duration}_{V2}$
(2) $\text{Intensity ratio} = \text{Intensity}_{V1}/\text{Intensity}_{V2}$
(3) $\text{Pitch ratio} = |39{,}86 \log_{10}(f0_{V2}/f0_{V1})|$.

## 3. RESULTS

The data were analysed using linear mixed models in R. Models were built up in stepwise fashion, using Likelihood ratio tests to determine best fit. The Tukey HSD procedure was used for post-tests on interaction effects.

### 3.1. Pitch ratio

For the pitch ratio, the best fitting model consisted of the random effects of age, context (whether disyllables were single utterances or from a multi-word context), participant identity and age crossed with participant identity. The fixed effects were age, group (CI or NH), status (babble or word) and group crossed with status. The fixed effects evidenced development over age *(E*=0.04, *S.E.=0*.15, *t*=2.49, *p*=.015*)* and that groups differed in their use of the pitch ratio *(E*=1.19*, S.E.*=0.23*, t*=5.27*, p*<.001*)*, as the NH disyllables had bigger semitone ratios compared to those from the CI group. Although the fixed effect of status significantly improved the model, it did not reach significance *(E*=-0.23, *S.E.*=0.17, *t* =-1.39*, p*=0.17*)*. There was a significant interaction between group and status *(E*=-0.51*, S.E.*=0.21*, t*=-2.43*, p*=.02*)*, which was examined through post-tests. These indicated that the NH infants had significantly smaller pitch ratios on babbles than on words (*E*=-0.74*, S.E.*=0.16*, z*=-4.68*, p*<.001). This was not the case for the CI group (*E*=-0.23*, S.E.*=0.17*, z*=-1.39*, p*=0.49). Whilst the NH group increased the pitch ratio on lexical disyllables, the CI group did so to a smaller degree (see Table 1).

**Figures 1 and 2:** Visualisation of the pitch ratios. Legend: Shaded area = confidence interval

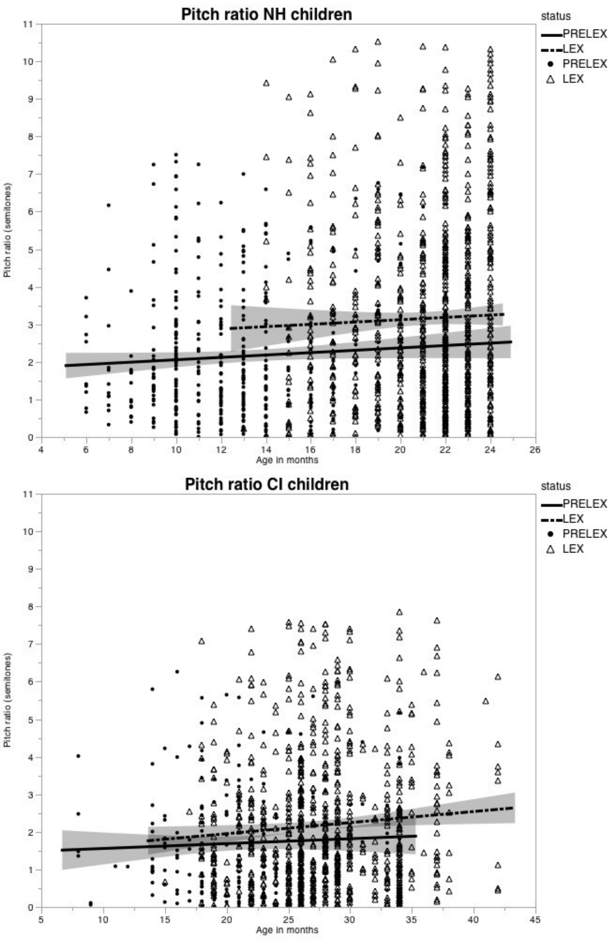


**Table 1**: The means and standard deviations for the ratios of the three acoustical cues.

| | Prelexical | | | Lexical | | |
|---|---|---|---|---|---|---|
| | Semitone | Intensity | Duration | Semitone | Intensity | Duration |
| CI | 1.73 (1.4) | 0.99 (0.08) | 0.73 (0.39) | 2.17 (1.81) | 1.02 (0.09) | 0.98 (0.49) |
| NH | 2.17 (1.73) | 1 (0.08) | 0.71 (0.89) | 3.16 (2.5) | 1.03 (0.08) | 0.91 (0.47) |

### 3.2. Intensity ratio

The best-fitting fitting model for the intensity ratio had age, participant identity and utterance context in the random effects. The fixed effects of age, group and status were needed for the best-fitting model. The effect of age *(E*=0.001*, S.E.*=0.001, *t*=2.37*, p*=.02*)* indicated that the use of the intensity ratio slightly increased over time. The NH group made slightly larger intensity differences overall between syllables *(E*=0.02*, S.E.*= 0.01*, t* = 2.37*, p* = .03*);* the main effect of status *(E*= -0.02*, S.E.*= 0.004*, t* = -3.17*, p* = .002*)* suggested that the intensity ratio was smaller for prelexical material in both groups.

### 3.3. Duration ratio

The best model for the duration ratio included age, participant, utterance context and age crossed with participant in the random effects. The following fixed effects were needed in the model: age, group, status and group crossed with status. There was a significant effect of age on duration ratio *(E*=0.01*, S.E.*=0.003, *t*=4.26*, p*<.001*),* indicating increasing duration ratios with age. Status was a significant main effect, as lexical disyllables had larger ratios in both groups *(E*=-0.16*, S.E.*=0.04*, t*=-4.35*, p*<.001*)*. There was also an interaction effect between groups and status (*E*=0.12*, S.E.*=0.05*, t*=2.53*, p*=.01). The post-test on the crossed effect showed that only the CI group made a significantly smaller duration ratio on prelexical utterances compared to lexical utterances (*E*=-016.*, S.E.*=0.04*, z*=-4.35*, p*<.001).

## 4. DISCUSSION

This study investigated the production of three acoustic cues to word stress, f0, intensity and duration, by NH and CI infants. We were interested in a developmental comparison of the two groups, and the influence of first words on acoustic cues.

For the pitch ratio, bigger differences in the NH group were found regardless of lexical status, and a further boost in ratio was seen once children started to produce words. Not only were the ratios of the CI children smaller by almost a semitone, but the marked increase in differentiation between stressed and unstressed syllables once a vocabulary started to emerge was not seen in this group, although there was a trend in the right direction. Given the limitations in F0 processing of the implant[10], it is not surprising that this seems to be an area of difficulty. Nevertheless, it is encouraging that the pitch ratios of the CI group follow the same direction as those of the NH group: a ratio above 1 indicates a trochaic pattern, which is the most frequent pattern in Dutch[23].

The intensity ratios of the NH group were also larger than those of the CI group irrespective of status. In both groups there was a smaller ratio for babbles. Note that for babbles, the ratio of the CI group indicated higher intensity on the second syllable, in opposition to the dominant trochaic pattern, which was not the case for the NH group. This may therefore indicate a locus of weakness; but once children with CI started to produce words, the intensity ratio not only increased, but also conformed to the trochaic pattern. Again, given the literature on the perception of intensity [19][21], difficulties with this cue are not surprising.

Both groups increased their duration ratios in words, but this was only significant for the CI group. It is however not clear whether this is truly an effect of learning to produce word stress, or whether utterance final lengthening [15] is at play, as the ratios indicated the second syllable was longer for both groups and in both types of disyllables, unlike the expected pattern for trochaic items [23]. Presently we are not able to unpeel the influences of different phrasal and prosodic levels on durations in the spontaneous speech of the participants, but it is hoped that future analyses of disyllables from multiword contexts may shed further light. The rest of this discussion therefore only concerns pitch and intensity ratios.

In summary, when comparing the two groups, the overarching pattern is that children with CI make the same use of acoustic cues as the NH group, but that differences between stressed and unstressed syllables are reduced, as evidenced by smaller ratios. This counters earlier reports of atypical stress [16], and is in accord with Hide [12], who reported correct word stress in 6 year old children with CI, in spite of smaller differentiation in acoustic cues. Unlike Hide's study, we do not yet know whether the children in our study are making enough use of the cues to produce perceptible stress differences, as perceptual ratings will be needed for this. The children in the present study, like the majority of the children in Hide's [12] study, were implanted before the age of 2, and it is of interest whether the early intervention has prevented some of the problems with stress perception and production reported in children who received their implants later.

In developmental terms, cues become more pronounced in words, with a striking jump to bigger pitch ratios in the NH group. The CI groups also boosts their ratios in words, but they do not manage this to the same extent as the NH group. Clearly, the fact that there is a communicative drive and an adult target behind children's first words engenders greater speech clarity, i.e. the more pronounced differences we are seeing. In addition, vocabulary development itself might be a force behind the crystallization of prosodic cues: analogous to the segmental domain, where Edwards et al.[11], amongst others, have described the dependence of phonetic learning on vocabulary development; for an argument that in speech perception, the acquisition of words may have a similar effect on categorical and distributional learning, see Werker et al. [26].

This possibility raises the question whether the less pronounced development in the words of children with CI is a simple delay which will resolve itself given more time, or whether they are failing to abstract something essential from the ambient input. Before implantation, the CI group also had extremely impoverished or almost absent input. They therefore started to speak later, and their phonetics may merely reflect this late start in perception and production. Furthermore, because they begin to speak later, at a more advanced cognitive stage, children with CI often start to produce first words and babbles at the same time, unlike NH children [20], which was the case for several children in our study, and may further have disrupted prosodic development.

It is unknown how the profile of dampened phonetic cues to word stress is reflected in the representations and linguistic processing of these children. Werker and Tees [26] have proposed a cascading model of language development, where disruption of the acquisition of earlier, lower-level processes can affect later, more complex language abilities. It may therefore be fruitful to trace the relationships between phonetic development and higher-level linguistic abilities in this population and NH children.

## 5. REFERENCES


[1] Ballard, D., Djaja, D., Arciuli, J., James, D., & van Doorn, J. 2012. Developmental Trajectory for Production of Prosody: Lexical Stress Contrastivity in Children Ages 3 to 7 Years and in Adults. Journal of Speech, *Language, and Hearing Research*, 55, 1822-1835.

[2] Beckman, M., Munson, B., & Edwards, J. 2004. Vocabulary growth and developmental expansion of types of phonological knowledge. *9th conference on Laboratory Phonology*.

[3] 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.

[4] Chapman, K., Hardin-Jones, M., Schulte, J., Halter, K. 2001. Vocal development of 9-month-old babies with cleft palate. *Journal of Speech, Language, and Hearing Research*, 44, 1268-1283.

[5] Chatterjee, M., Peng, S-C. 2008. Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hearing Research*, 235, 143-156.

[6] Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.

[7] Davis, B., MacNeilage, P., Matyear, C., Powell, J. 2000. Prosodic correlates of stress in babbling: An acoustical study. *Child Development*, 71, 1258-1270.

[8] 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.

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

[10] 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.

[11] Edwards, J., Munson, B., & Beckman, M. E. 2011. Lexicon-phonology relationships and dynamics of early language development: A commentary on Stoel-Gammon's relationships between lexical and phonological development in young children. *Journal of Child Language*, 38, 35–40.

[12] Hide, O. 2013. *Acoustic features of speech by young cochlear implant users. A comparison with normal-hearing and hearing-aided age mates. (PhD*) Antwerp: University of Antwerp.

[13] Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., Lyons, T. 1991. Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27(2), 236–248.

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

[15] Lehiste, I. 1970. Temporal organization of higher-level linguistic units. *Journal of the Acoustical Society of America, 48*(1), 111.

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

[17] MacWhinney, B. 2013. CLAN (Version 04-Nov-2013 11:00).

[18] Mattys, S., Jusczyk, P., Luce, P., Morgan, J. 1999. Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, 38, 465-494.

[19] Meister, H. 2011. Processing Prosodic Cues with Cochlear Implants. *Sprache-Stimme-Gehor*, 35(3), 99-104.

[20] Moreno-Torres, I. 2014. The emergence of productive speech and language in Spanishlearning paediatric cochlear implant users. *Journal of Child Language*, 41(3), 575-599.

[21] O'Halpin, R. 2010. The perception and production of stress and intonation by children with cochlear implants. (PhD), London: University College London.

[22] Peng, S-C., Tomblin, B., Turner, C. 2008. Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *Ear & Hearing*, 29, 336-351.

[23] Sluijter, A., van Heuven, V. 1996. Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proceedings of the International Congress of Phonetic Sciences*.

[24] 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.

[25] Werker, J., Tees, R. 2005. Speech Perception as a Window for Understanding Plasticity and Commitment, Language Systems of the Brain. *Developmental psychobiology*, 46, 233-251.

[26] Werker, J., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception?. *Current Directions in Psychological Science*, 21(4), 221-226.

[27] 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.