

Speech intelligibility of children with an auditory brainstem implant: a triple-case study

Authors

Jolien Faes

Computational Linguistics & Psycholinguistics (CLiPS) Research Center, University of Antwerp, Belgium

Sven De Maeyer

Department of Training and Education Sciences, Faculty of Social Sciences, University of Antwerp, Belgium

Steven Gillis

Computational Linguistics & Psycholinguistics (CLiPS) Research Center, University of Antwerp, Belgium

Correspondance:

jolien.faes@uantwerpen.be

Funding

This work was supported by the Research Foundation in Flanders (FWO) under grant [12Q6318N].

Acknowledgements

This work was supported by the Research Foundation in Flanders (FWO) [grant 12Q6318N]. We would like to thank the children and their parents for participating in this study, and N. Boonen, I. Molemans, R. van den Berg, L. Van Severen and K. Schauwers for the collection of most video-recordings. We thank L. Swinnen for the RIZIV information about the number of pediatric ABI implantations in Belgium.

Declaration of interest

There was no conflict of interest reported.

Abstract

Auditory brainstem implantation (ABI) is a relative recent development in paediatric hearing restoration. Consequently, young-implanted children's productive language has not received much attention. This study investigated speech intelligibility of children with ABI (N=3) in comparison to children with cochlear implants (CI) and children with typical hearing (TH). Spontaneous speech samples were recorded from children representing the three groups matched on cumulative vocabulary level. Untrained listeners (N=101) rated the intelligibility of one-word utterances on a continuous scale and transcribed each utterance. The rating task yielded a numerical score between 0 and 100, and similarities and differences between the listeners' transcriptions were captured by a relative entropy score. The speech intelligibility of children with CI and children with TH was similar. Speech intelligibility of children with ABI was well below that of the children with CI and TH. But whereas one child with ABI's intelligibility approached that of the control groups with increasing lexicon size, the intelligibility of the two other children with ABI did not develop in a similar direction. Overall, speech intelligibility was only moderate within all groups of children, with quite low rating scores and considerable differences in the listeners' transcription, resulting in high relative entropy scores.

Introduction

Pediatric hearing restoration of severe-to-profound hearing loss has long been restricted to sensorineural hearing deficits situated within the cochlea. With a cochlear implant (CI) an electrode array is inserted into the cochlea bypassing absent or malformed hair cells of the cochlea and directly stimulating the auditory nerve. Since 2001, also other inner ear pathologies causing pediatric severe-to-profound hearing loss became treatable by extending the use of an auditory brainstem implant (ABI) from adults to children (Colletti et al., 2001).

An ABI is appropriate when the hearing loss results from, for instance, the absence of the auditory nerves, cochlear ossification, or cochlear malformation, in which cases a CI cannot be implanted. An ABI is also used as an alternative option when children's speech and language with CI is not developing as expected (Batuk et al., 2020). An ABI directly stimulates the cochlear nucleus of the brainstem, bypassing the cochlea and the auditory nerve (Puram & Lee, 2015). In that sense, there is an important difference between auditory stimulation by a CI versus an ABI. Whereas the CI stimulates the spiral ganglion – known to be highly tonotopically organized -, the ABI stimulates the neural pathways of the brainstem, which are identified as unpredictable and of which the tonotopic organization is as yet unknown (Wong et al., 2019).

Pediatric ABI implantation is a relatively recent development compared to the pediatric CI. For children, the first CIs were implanted in the early 1980s, whereas the first ABIs was only implanted at the beginning of this century. Even though fewer children are eligible for an ABI than for a CI (Kaplan et al., 2015), a growing number of children is receiving ABIs worldwide, especially in Europe and the US (amongst many others: Colletti et al., 2014; Sennaroglu, Colletti, et al., 2016; Wilkinson et al., 2017). Consequently, the literature on children with ABI is still rather limited and concerns small cohorts of participants. Moreover, scientific reports mainly focus on children's evolving speech perception skills but far less is known about their spontaneous speech production and communication skills.

Children with ABI

Research has shown that the ABI provides sound awareness for congenitally severe-to-profoundly deaf children. Aided hearing levels (PTA, pure tone average) with the implant can be expected between 30 and 60 dB HL (decibel hearing level) (Sennaroglu, Colletti, et al., 2016; e.g. Teagle et al., 2018; Wilkinson et al., 2017; Yucel et al., 2015). Functionally

speaking, children with ABI are expected to be able to discriminate and identify sounds and phonetic contrasts. Speech perception (and therefore also speech production skills) are more developed in children with lower hearing thresholds with ABI (Sennaroglu, Sennaroglu, et al., 2016; Yucel et al., 2015), who were implanted earlier (Aslan et al., 2020), and without additional disabilities (Colletti et al., 2014; van der Straaten et al., 2019). Children with ABI meeting these criteria are in the best position to develop open set speech perception. Most of them can reach CAP scores of at least five on a seven-point scale (Categories of Auditory Performance, Archbold et al. (1995)), indicating that they are able to understand simple sentences without lip-reading (Colletti et al., 2014; Sennaroglu, Colletti, et al., 2016). However, also a considerable amount of individual variation has been reported (van der Straaten et al., 2019).

ABI also appears to have a beneficial effect on children's speech production skills, although the development is very slow (Aslan et al., 2020; Eisenberg et al., 2018; Faes & Gillis, 2019a, 2019b, 2020; Fernandes et al., 2020; Teagle et al., 2018; van der Straaten et al., 2019). The best results appear in children with ABI with the same characteristics as for speech perception: no additional disabilities, low hearing thresholds after implantation and early implantation (van der Straaten et al., 2019). With extended device use, these children with ABI start vocalizing, babble and – later on – produce words and sentences (Bayazit et al., 2014; Faes et al., 2019; Faes & Gillis, 2019a, 2020; Puram & Lee, 2015). They clearly expand their lexicon sizes with increasing hearing experience (Faes & Gillis, 2019b), start to use basic word patterns (Eisenberg et al., 2018; Faes & Gillis, 2020), and produce mono- and disyllabic words and language ambient phonemes after two to three years of device use (Eisenberg et al., 2018; Faes & Gillis, 2020, 2021; Teagle et al., 2018). However, the accuracy of their speech production is fairly limited at the phoneme level (Eisenberg et al., 2018; Faes & Gillis, 2021; Teagle et al., 2018) and at the word level (Faes & Gillis, 2020).

From the available studies, it can safely be concluded that ABI implantation has a clear effect on children's spoken language development, especially for early implanted children without additional disabilities and with low aided hearing thresholds. However, progress is slow and stays well below the expected progress in children with CI and children with typical hearing (TH). The better performing children with ABI have expressive language skills that can be situated between those of children with CI with additional disabilities and children with CI without additional disabilities, even after five or six years of device use (van der Straaten et al., 2019). Also for lexical development, children with ABI's vocabulary sizes lie well below those of children with CI and children with TH without additional disabilities and with the same amount of hearing experience (Faes & Gillis, 2019b). The same holds for phonological complexity in production and for word production accuracy: children with ABI's performance can sometimes be situated in the lower ranges of the 95% intervals of these control groups, but more often outside of these confidence intervals, even after several years of device use and even when vocabulary sizes were matched (Faes & Gillis, 2020).

The literature suggests that speech production development is very slow in children with ABI. Their language and speech are less advanced as compared to children with CI and children with TH with the same chronological age or hearing age (device experience). It appears to take several years of device use for these children to produce ambient language phonemes and first words. Hence, the overarching question turns up how their intelligibility develops. Speech intelligibility offers a general view of children's speech production skills since intelligible speech production involves the incorporation of all linguistic skills at once when speaking. According to Yucel et al. (2015), children with ABI's speech intelligibility is a weakness. Reaching intelligibility is an even more protracted process in comparison with children with typical hearing and children with a cochlear implant. The aim of the present

paper is to compare intelligibility in three cases with ABI matched with those two other groups at specific linguistic levels.

Speech intelligibility and its metrics

In the present study, speech intelligibility is conceptualised as the extent to which a listener can correctly recover particular elements (e.g., phonemes, words) in an acoustic signal generated by a speaker (V. Freeman, D. B. Pisoni, et al., 2017; van Heuven, 2008; Whitehill & Ciocca, 2000). As such intelligibility can be distinguished from comprehensibility. The latter refers to the process on the side of the listener of reconstructing the intended meaning or the message conveyed by the speaker's acoustic signal. In order to elucidate the difference between intelligibility and comprehensibility, suppose that the perfectly grammatical sentence "Colorless green ideas sleep furiously" (Chomsky, 1957) is read to a group of speakers of English. When asked to transcribe the sentence, i.e., to literally write down the sentence, they will probably be able to write down the words comprising the sentence. That is, the sentence is *intelligible*. However, the meaning of that sentence, the intended message of the speaker, is at least quite opaque, not to say that the sentence is *incomprehensible*.

Speech intelligibility is an important yardstick in speech language development, as becoming intelligible to others is seen as an important objective in child language development. A child who is intelligible for unfamiliar listeners is believed to have acquired all aspects of linguistic and cognitive skills, speech perception and speech production required for successful communication (S. Freeman et al., 2017). By extension, children's level of speech intelligibility is often used as a clinical tool: it is used to measure the progress of therapy and a good indicator for directing children to speech and language therapy if intelligibility is considered to be too low relative to age norms (Chin et al., 2012; Gordon-Brannan & Hodson, 2000). Typically developing children's speech is intelligible for

unfamiliar listeners approximately by the age of four (Baudonck et al., 2009; Chin et al., 2003; Flipsen & Colvard, 2006; Hustad et al., 2020). Children with CI typically score lower in intelligibility tests as compared to hearing age-mates (Chin et al., 2003; S. Freeman et al., 2017). Even after approximately seven years of device use they do not reach the same intelligibility scores as normally hearing children in a sentence imitation task (Chin & Kuhns, 2014).

Speech intelligibility is often measured using the *Speech Intelligibility Ratings* or SIR (Cox & McDaniel, 1989), used by e.g. Calmels et al. (2004), De Raeve (2010), Lejeune and Demanez (2006) and Toe and Paatsch (2013) in children with CI. This ordinal scale ranges from the child using only prerecognizable words in spoken language (level 1 on the SIR scale) to the highest level (level 5), meaning that the child's connected speech is intelligible to all listeners in everyday contexts. One disadvantage of the SIR is that its ordinally ranked categories are fairly coarse. As a net result, early implanted children with CI reach the upper limit of the SIR already after three years of device use (De Raeve, 2010), even though there are still unintelligible parts in their speech (Miller, 2013). Other numeric ratings scales have been used in the literature thus far (e.g. AlSanosi & Hassan, 2014; Habib et al., 2010; Tseng et al., 2011). For instance, a seven-point scale with only the first and the last position on the scale identified as being completely unintelligible and completely intelligible (Habib et al., 2010; Peng et al., 2004). These rating scales can be used with various types of speech productions, including imitated speech and spontaneous speech.

In addition to rating scales, also so-called *objective* ratings (Hustad et al., 2020) have been used in the literature, mostly operationalized as transcription tasks. In other words, listeners transcribe children's utterances (henceforth stimuli) orthographically or phonetically. When the stimuli are derived from a predefined set of words or sentences (e.g., in an imitation task, in a picture-naming task, in a reading task), a comparison between the listener's transcription

and the target can straightforwardly be made, resulting in a number of correctly identified targets and an overall percentage of intelligibility. But when the stimuli originate from children's spontaneous speech productions, a comparison with the target is difficult, if not impossible, since only the child knows what the actual target was. Hence, when the intelligibility of spontaneous speech is assessed, a transcription can only be compared with an unknown target and consequently a straightforward correct/wrong evaluation is impossible. For transcription tasks without a predetermined target, several alternatives for calculating intelligibility have been proposed. One option is to calculate the number of (un)intelligible syllables or words identified by the listeners as an index of intelligibility (Flipsen & Colvard, 2006; Lagerberg et al., 2014; Strömbergsson et al., 2020). Another option is to use multiple transcriptions of the same sample of spontaneous speech and to calculate the relative entropy of the listener's transcriptions. The underlying assumption is that the more diverse listener's transcripts, the higher the relative entropy and thus the lower the child's intelligibility. Relative entropy was, for instance, used in linguistic studies on the mutual intelligibility of related languages, such as Swedish and Danish (Frinsel et al., 2015; Moberg et al., 2007). Using this relative entropy metric, Boonen (2020) showed that children with CI's speech intelligibility is significantly lower than that of children with TH at seven years of age.

Speech intelligibility in children with ABI

Most children with ABI reach level 1 on the SIR after approximately one year of device use, meaning that they produce prelexical vocalizations or, in SIR's terminology, prerecognizable words and used their voice as an attention getting device (van der Straaten et al., 2019). The children with the highest speech intelligibility reach level 3 or 4, i.e., their speech is intelligible for an experienced listener with or without lip-reading (Aslan et al., 2020; Sennaroglu, Sennaroglu, et al., 2016; van der Straaten et al., 2019). These children are

implanted before their fifth birthday, have relatively low aided hearing thresholds and no additional disabilities, but it takes them five to six years to reach these intelligibility scores.

However, in the literature on speech intelligibility of children with ABI, the procedures for obtaining speech intelligibility scores are not always well articulated and remain rather vague. For instance, Aslan et al. (2020) judged speech intelligibility using the SIR based on children's connected speech, but it is not indicated which amount of connected speech was evaluated and only one clinician evaluated the children's speech. Sennaroglu, Sennaroglu, et al. (2016) and van der Straaten et al. (2019) also assessed speech intelligibility of ABI children with the SIR, but they do not mention how many judges were involved, nor on how many words or if the speech was produced spontaneously or not. In the present study, one hundred untrained listeners judged spontaneous isolated word productions of children with ABI in a rating task and in a transcription task.

In the present study, the speech intelligibility of children with ABI was assessed in comparison with children with CI and children with TH. In principle, there were several options for matching the various study groups. They could be matched on their chronological age, on their hearing age, or on a language related measure such as mean length of utterance (Brown, 1973) or vocabulary size (see e.g. Faes & Gillis, 2016 for a more elaborated discussion). Using chronological age as a yardstick for comparing children's intelligibility across different hearing conditions was discarded since it would have led to a comparison of children's spoken language performance at vastly different ages. More specifically, children with ABI are typically implanted after the age of two, while children with CI are commonly implanted before their first birthday. This means that comparing the children at the age of four, for instance, implies that the children with TH have four years of hearing experience, as compared to approximately three years for the children with CI and only two years for the children with ABI. Hence, the differences between their hearing experiences may have led to

differences in their speech and language development and different intelligibility. In order to take into account the prolonged period of auditory deprivation of children with CI and ABI, hearing age – i.e. the length of device use –, has often been used as an alternative in the CI literature, in order to compare these children to children with TH (e.g. Caselli et al., 2012; Ertmer & Goffman, 2011; Schramm et al., 2010). But hearing age is also a time-based measure, and, hence, also subject to chronological age-related differences in children’s speech motor control (see e.g. Faes & Gillis, 2016). Therefore, the use of more language intrinsic measures has been advocated in the literature (Faes & Gillis, 2016; Santos & Sosa, 2015). One of the measures related to linguistic maturation or “language age”, is vocabulary size. Research has shown that lexical development and phonological development are closely related in children with TH (among others: Sosa & Stoel-Gammon, 2006; Stoel-Gammon, 2011; van den Berg, 2012) and children with CI (Faes & Gillis, 2016; Nicholson et al., 2015; Reidy et al., 2015). Since speech intelligibility is also, but not solely, linked to children’s speech production accuracy and thus phonological development (Ingram, 2002), matching the groups of children with ABI, with CI and with typical hearing on their level of lexical development was also adopted in the present study.

Research aims

The research question addressed in the present study is as follows: How intelligible are the spontaneous speech productions of children with ABI in comparison to children with TH and children with CI matched on lexical age? For this purpose, a longitudinal triple-case report of three children with ABI is presented– in comparison to peers with CI and TH. Nagels et al. (2020) highlighted the importance of tracking individual patterns of language development for heterogeneous clinical groups, such as children with CI. As can be derived from the literature study above, children with ABI are a highly diversified group as well. Therefore, the

adopted case-study approach allows a fine-grained study of individual patterns in the three children with ABI in this study.

Isolated single words were selected from spontaneous speech samples for the speech intelligibility measurements. One hundred individuals not familiar with the children rated each speech sample on a continuous scale and also transcribed each word. The ratings and the transcriptions of the samples of the children with ABI, TH and CI were analyzed. For the latter, relative entropy was used to investigate the amount of consistency in the judges' transcriptions. Each child with ABI was matched to peers with TH and peers with CI with similar levels of lexical development.

Method

This study reports on a listener experiment aimed to investigate the speech intelligibility of three children with ABI in comparison to children with CI and children with TH with similar vocabulary levels. Three consecutive steps were taken in setting up the experiment, which will be described in the present section:

- (1) Longitudinal data collection of participants with ABI and their matching peers with CI and TH;
- (2) Experimental setup: selection of suitable stimuli from the data collected in (1);
- (3) Actual experiment: procedure and participants.

After the description of these three steps, the procedures of data processing and statistical analyses will be elaborated on.

Longitudinal data collection

Three children with ABI participated in this study and two control groups comprising children with cochlear implants (CI) and children with typical hearing (TH) were included. This study was approved by the Ethical Committee for Social and Human Sciences of the XXX.

Children with ABI

The pool of children with ABI implanted before the age of five is still very limited in Belgium. According to statistics of the RIZIV (the Belgian national institute for Health and Disability Insurance), only eight children received an ABI before their fifth birthday between 2015 and the end of 2019. Two criteria restricted the number of children eligible for participation in the present study. First, only Dutch-speaking children were included into the study. Since Belgium has three regions, each with its own official language, only children living in the northern, Dutch-speaking part of Belgium (Flanders) were eligible. Second, also children with reported developmental or health problems were excluded from the data collection. These criteria restricted the number of participants to three cases, henceforth referred to as ABI1, ABI2 and ABI3.

ABI1 and ABI2 were born with a sensorineural profound hearing loss as a result of the absence of the auditory nerves. They received an ABI at age 2;00 (years;months) and 2;01 respectively. Their Pure Tone Average (PTA) hearing thresholds improved from respectively 120 dB HL and 116 dB HL before implantation to 37.5 dB HL and 43 dB HL two years after surgery, according to their medical records. In both children, 9 out of 12 electrodes were activated approximately one month after the surgery. ABI1 received a second ABI at age 4;09. ABI3 was first implanted with a CI (at age 0;08), after a diagnosis of auditory neuropathy. Even though the child's PTA improved from 95 dB HL (in the better ear) to 33 dB HL after CI implantation, there was only a limited effect on speech and language

development. Therefore, the child received a contralateral ABI at age 4;00. The implant was fitted two months after the surgery and all electrodes were activated.

The children with ABI were raised orally in Dutch, with support of Flemish Sign Language. The children, data were collected longitudinally and monthly as part of a larger research project on their speech and language development. Data collection started one year after implantation for ABI1, two years after implantation for ABI2 and immediately after implantation for ABI3 and went on for about two years in all three cases.

Please insert Table 1 over here.

Control groups

A first control group consisted of nine children with CI. These children received a CI (mean age 1;00, SD = 0;05) because of a congenital profound deafness with a mean PTA of 112.56 dB HL before implantation. The mean PTA improved to 32.22 dB HL (SD = 7.11) at the children's second birthday. Six children received a second CI later on (see Table 2). All children were raised in oral Dutch, with a limited number of lexical signs. Data collection started immediately after implantation, with a monthly follow-up up to 30 months after implantation. In Table 2, individual data for the children with CI are presented.

A second control group consisted of children with typical hearing (TH). As part of a larger research project, 30 children were followed longitudinally and monthly between ages 0;06 and 2;00.

Please insert Table 2 over here.

Experimental setup: selection of suitable stimuli

Matching of children with ABI and control groups

The material from which the stimuli for the current study were selected consisted of longitudinal monthly video recordings of the children with ABI, CI and TH. The recordings were made as part of larger longitudinal research projects on spontaneous language and speech development of the three groups of children. They comprised everyday spontaneous interactions between the child and his/her caregiver(s) captured at the children's homes. Each recording lasted on average approximately one hour. All children's utterances were transcribed orthographically in CLAN according to the CHAT conventions (MacWhinney, 2000).

In order to compare the three children with ABI to the children with CI and TH relative to their cumulative vocabulary, a cumulative vocabulary count was computed for each individual child in the three groups of children. This means that the number of unique word types in the first recording was incremented each time with the new word forms in the following monthly recordings. In this way the increase of each child's vocabulary was tabulated (Faes & Gillis, 2019b).

Given the vocabulary counts, the next step in the selection process of the experimental stimuli consisted of matching the levels of vocabulary development of the three groups of children as closely as possible. Since the data collection was longitudinal, no preset level of vocabulary could be used. For each of the participants (ABI, CI and TH), a graph was drawn with the cumulative vocabulary relative to the child's hearing age (in months). An overview is given in Figure S1 in the appendix.

First, the three cases with ABI were inspected. As can be derived from Figure S1, there is a substantial amount of difference in the three children with ABI. On the one hand, this variation is inherent to interindividual differences in children acquiring language (Kidd & Donnelly, 2020), but on the other hand, also inherent to the schedules of the present data collection (e.g., for the ABI children data collection started immediately after implantation in

one case, one year or two years after ABI implantation). Five levels of lexical development were selected as a function of data availability. The five levels of lexical development were: (1) less than 50 word types, (2) ca. 100 word types, (3) ca. 200 word types, (4) ca. 350 word types, and (5) more than 500 word types. Henceforth, we will label these levels as level (1) to level (5). Three data points were chosen for ABI1 and ABI2 and two data points for ABI3. So, a total of 8 ABI recordings were selected for this study. In Table 1, the different data recordings of all children are presented, with their corresponding ages, hearing ages and cumulative vocabulary sizes. For ABI1, recordings were selected at lexical level (1) less than 50 word types, level (2) ca. 100 word types, and level (4) ca. 350 word types. For ABI2, recordings were selected for lexical level (3) ca. 200 word types, level (4) ca. 350 word types, and level (5) more than 500 word types. For ABI3, recordings were selected for lexical level (2) ca. 100 word types and level (4) ca. 350 word types.

Second, the matching with the control groups was performed based on these five levels of lexical development. Since the discrepancy between the number of children with ABI (three cases) and the number of CI and TH participants (N=9 and N=30 respectively), a random selection of available data in the control groups was made. For each level of lexical development, three recordings of children with TH and three recordings of children with CI were matched. This matching was random, but again as a function of data availability. The individual cumulative vocabulary counts of all children, as presented in Figure S1, were used.

For children with CI, recordings were available for all five levels of lexical development. At level (4) ca. 350 word types, five instead of three CI recordings were selected in function of data availability. In total, 17 CI recordings were selected. For children with TH, no data were available at the highest level of lexical development, i.e., level (5) more than 500 word types. This resulted in a total of 12 TH recordings: three recordings for level (1) less than 50

word types, three recordings for level (2) ca. 100 word types, three recordings for level (3) ca. 200 word types, and three recordings for level (4) ca. 350 word types.

Selection of suitable stimuli

Of all selected recordings, only one-word utterances with no background noise, crosstalk, and the like were eligible for further use. Of this subset, the monosyllabic and disyllabic words were filtered out. From these, 10 utterances per recording were randomly chosen. This resulted in a total of 370 utterances, with 80 ABI utterances (8 ABI recording x 10 utterances), 170 CI utterances (17 CI recording x 10 utterances) and 120 TH utterances (12 TH recording x 10 utterances). For two children with CI, there were too little one-word utterances. In those cases, a two-word utterance was chosen, with the first word being an article. This was the case in only 5 out of the 370 stimuli used in the present study.

The actual experiment: procedure and participants

The 370 selected stimuli were divided into five experimental series of utterances, each containing 74 stimuli. The process of compiling the five series was basically random with the constraint that each series comprised a proportional number of ABI, CI and TH samples: two of the 10 selected stimuli of the recording of each child were randomly selected, resulting in 2 x 8 ABI stimuli, 2 x 12 TH stimuli and 2 x 17 CI stimuli.

All stimuli were entered into Qualtrics[®] (Qualtrics, Provo, UT). A total of 101 untrained listeners participated in the study, with a minimum of 20 listeners for each series. Listeners were randomly selected through snowball sampling starting from the personal acquaintances of the authors of the present paper. They were all native speakers of Belgian Dutch (mean age = 37 years, SD = 13 years), with no self-reported history of hearing loss and varying degrees of experience with children's language, but no experience with the speech of hearing-

impaired children. A non-parametric Wilcoxon test revealed no significant impact of the listeners' experience with child language on the outcomes of the rating scale and transcription task.

The participating listeners completed the experimental tasks at their own convenience in their home environments. They were instructed to wear earphones or headphones. Before the actual experiment started, instructions were presented on screen, and examples were given of the experimental tasks in order to ensure that the participant understood the instructions. Each participating listener completed one of the series of stimuli, which was randomly assigned. In addition, the order of the 74 stimuli was randomized upon each presentation, so that in principle each listener heard the stimuli comprising a series in a different order.

For each stimulus, listeners performed two tasks, represented in Figure 1: (1) indicate the intelligibility of the utterance, and (2) transcribe the utterance. For the first task, listeners judged the utterances' intelligibility by moving a slider on a 100-point scale, going from entirely unintelligible to entirely intelligible. Since the listeners were untrained, the SIR was not used in order not to complicate their task. For the transcription task, listeners were instructed to write down existing standard Dutch words, i.e., words that they thought the child produced or was trying to produce. If they could not figure out which word the child intended, they were instructed to write the character 'x'.

Please insert Figure 1 over here.

Data analyses

The experiment resulted in a rating score (between 0 and 100) and a transcription for each of the stimuli. For the transcription task, the position of the slider was transformed into a natural number between 0 and 100 by Qualtrics, which was entered into the statistical analyses. The data of the transcription task consisted of the transcriptions of the participants. Relative

entropy was used as a measure of the consistency of the transcriptions between the participants. The underlying assumption was that if all the transcribers agreed on a particular transcription, then the child's word must have been very intelligible. But if all transcribers disagreed and/or used the symbol 'x' for denoting an unidentifiable word, then the child must have been very unintelligible. Relative entropy quantifies the degree of agreement between the transcribers. More specifically, entropy is a measure of chaos or disorganization in data, often used in information theory and also sporadically used in linguistic research to measure, for instance, the mutual intelligibility of languages (Frinsel et al., 2015; Moberg et al., 2007) or the intelligibility of children's language (Boonen, 2020).

Relative entropy was calculated per stimulus and indicates the degree of agreement between the listeners. Relative entropy is calculated according to the equation in (1), using Shannon's original entropy (Shannon, 1948) divided by the maximum entropy:

(1)

$$Entropy = \frac{-\sum_{i=0}^n (p_i \ln (p_i))}{\ln(N)}$$

with p_i = the probability of each transcription's occurrence; n = the total number of occurrences; and N = the number of listeners

A relative entropy score of 0 indicates complete correspondence over all transcriptions, thus indicating that all listeners agree on the transcription and hence indicating high intelligibility. A relative entropy score of 1 designates the opposite: none of the listeners gave the same transcription, thus indicating complete disagreement between the listeners and hence low intelligibility.

For the calculation of the relative entropy, the answer 'x', i.e. when a listener had no idea which word a child produced, was considered as a unique answer. Thus, for instance, when the transcriptions of three listeners were 'x', these were entered into the computation of entropy as 'x₁', 'x₂', and 'x₃'. The correlation of calculating the relative entropy with answer 'x' as unique answers and the relative entropy with all answers 'x' as one single answer was

0.94 ($p < 0.001$). The correlation between the relative entropy and the mean rating scores equaled -0.82 ($p < 0.001$). In other words, both measures are similarly sensible to the children's speech intelligibility. This observation was already made in the literature as well (e.g. Habib et al., 2010; Peng et al., 2004).

Statistical analyses

Given the design of the study, the resulting observations cannot be seen as independent observations. The rating scores for a certain stimulus are, on the one hand, nested within children and, on the other hand, nested within raters. In other words, a rating score depends on the rater and on the child heard by the rater. For the entropy measures the structure in the data is less complex, but still hierarchical: entropy scores are nested within individuals. To account for this complexity in the data multilevel models were used. These models consist of two parts: a random part that takes into account the variation and nesting of the data as described above, and a fixed part that models the predicted variables (Baayen, 2008).

The dependent variable was either the rating score or the relative entropy. Rating scores were converted to z-scores for the entire sample in order to catch the distribution of the scale. In this way, different rating behaviors between raters is controlled for. In order to normalize the skewed distribution, also relative entropy was log-transformed (\ln) (Baayen, 2008).

The children with TH were the reference category (i.e. intercept). The different grouping point categories (see paragraph *stimuli* and Table 1) were added as dummy variables in the model. Next dummy variables for the different children with ABI at these different grouping points, and a dummy variable for the CI data in interaction with the grouping point categories were added to the model as well. Random effects were child ID for the models estimating the relative entropy, and child ID, listener ID and stimulus ID for the models estimating the rating scores.

All statistical analyses were performed in R (R Core Team, 2013) with the package *lme4* (Bates et al., 2015). The predicted values of each model were used to resample the data by using the *predictInterval* function in R in the *merTools* package (Knowles & Frederick, 2020). This allowed creating a prediction interval around the fitted values of the model including the variation of children with CI and TH captured by the model. This resampling was done because there was no variation in the model for the children with ABI. The prediction interval was set at 90% and 10,000 resamples were taken. The distribution of the data resulting from this resampling procedure was plotted in the Result section. For the sake of convenience, the rating z-scores and the log-transformed relative entropy were reversed to their original scale in all figures.

Results

Intelligibility according to the rating scale

In Figure 2, the predicted median rating score (with 90% confidence intervals), scoring the intelligibility of the utterances produced by the three children with ABI in comparison with their peers with CI and TH, are plotted relative to lexical age, i.e., the cumulative number of word types. Comparisons of children with ABI with children with CI and TH were based on the number of cumulative word types in their lexicon: level (1) less than 50 word types (ABI1), level (2) ca. 100 word types (ABI1, ABI3), level (3) ca. 200 word types (ABI2), level (4) ca. 350 word types (ABI1, ABI2 and ABI3) and level (5) more than 500 word types (ABI2).

In the relevant (lexical age) time frame, the median rating scores of children with CI's and TH's utterances increased from approximately 40 to 65 on a 100-point scale. Thus, these children's intelligibility increased with lexical expansion.

With a small lexicon of less than 50 word types (level (1)), ABI1's utterances were rated at the same level as the utterances of children with CI and TH: the predicted median was slightly lower for ABI1 (score 37), but there was an overlap of the confidence intervals. With lexical expansion to level (2) ca. 100 and level (4) ca. 350 words, ABI1's utterances were rated systematically lower than these of children with CI and TH. ABI1's ratings remained between 30 and 40 on the 100-scale, whereas children with CI and TH showed an increase in the ratings scores to approximately 60.

ABI2's utterances, in contrast, were rated only slightly lower than these of children with CI and TH. At all lexicon sizes from ca. 200 to more than 500 word types (levels (3), (4) and 5)), ABI2 was approximating the control groups with matched cumulative vocabulary sizes. This child also showed increasing ratings with increasing lexical age.

With a vocabulary size of ca. 100 words (level (2)), ABI3's utterances are rated well below these of children with CI and TH, similar to ABI1's whose values were also rated below these of children with CI and TH, but the confidence intervals of ABI3 did not overlap with those of the CI and the TH children. As for ABI1, the lexical expansion to ca. 350 word types (level (4)) in ABI3 did not result in a considerable increase of the rating scores, so that the difference with children with CI and TH was maintained.

Please insert Figure 2 over here.

Intelligibility according to the transcription task: relative entropy

In Figure 3, relative entropy (predicted median and 90% confidence interval) for the three children with ABI, and the children with CI and TH, was plotted as a function of lexical age (number of word types). Relative entropy is a measure of uniformity or the lack of it in the transcriptions of the listeners in the transcription task. It is assumed that lower entropy scores are an index of more intelligible utterances. A relative entropy score of 0 indicates that all

listeners transcribed the utterance as the exact same Dutch word and, hence, it is assumed to be a completely intelligible utterance. A relative entropy score of 1 indicates the opposite: a complete unintelligible utterance, with different transcriptions from each listener.

For children with CI and TH, the predicted median relative entropy scores of their utterances progressed from approximately 0.70 with a lexicon size of less than 50 word types (level (1)) to approximately 0.60 with lexical expansion to 100 word types and more (levels (2) to (5)), so that their utterances can be assumed to become more intelligible with increasing lexicon size.

With a small cumulative vocabulary size of less than 50 word types (level (1)), ABI1's utterances had similar predicted median relative entropy scores than these of the children with CI and TH. However, with lexical expansion, the relative entropy of ABI1 hardly changed (ca. 0.72 in the entire period), so that the difference with children with CI and TH became apparent at lexicon sizes of 100 and 350 word types (levels (2) and (4)).

For ABI2, the predicted median relative entropy was only slightly higher than that of children with CI and TH at a cumulative vocabulary size of ca. 200 word types (level (3)). With lexical expansion to level (4) ca. 350 word types, ABI2's relative entropy estimates remained about the same (0.68), whereas there already was a decrease of the relative entropy for children with CI and TH. At a lexicon size of more than 500 words (level (5)), ABI2 also showed this decrease, resulting in similar relative entropy values at this point (0.65 for ABI2 and 0.63 for children with CI).

Finally, ABI3's utterances had predicted median relative entropy scores of 0.72 at lexicon sizes of ca. 100 (level (2)) and ca. 350 word types (level (4)). This was considerably higher than the values of children with CI and TH with a comparable lexicon size. Moreover, ABI3's confidence intervals did not overlap at all with children with TH, and only in part with children with CI at a lexicon size of ca. 350 word types (level (4)).

Please insert Figure 3 over here.

Discussion

This study investigated the speech intelligibility of three children with ABI in comparison to children with CI and children with TH matched on lexicon size. Speech intelligibility is postulated to be the most encompassing spoken language skill to develop for children with ABI (Aslan et al., 2020). Overall, the results of our large-scale listener experiment revealed that the speech intelligibility of children with CI was similar to that of children with TH, even though a slight advantage for the children with TH in median ratings and relative entropy scores appeared with increasing vocabulary sizes. This is in line with the literature on CI and TH speech intelligibility (Chin et al., 2012; Flipsen & Colvard, 2006; Grandon et al., 2020). For the three children with ABI, speech intelligibility is lower in comparison to their peers with CI and TH matched on lexicon size. Moreover, the triple-case study indicated interindividual differences between the children with ABI. Whereas ABI2 appeared to approach similar levels of intelligibility as children with CI and TH when matched on lexicon size, the other two children with ABI did not (ABI1 and ABI3).

Within the acquisition of the first 50 word types, ABI1 was as intelligible as children with CI and TH. It is well known that children's early words are often produced quite accurately in TH populations (Ferguson & Farwell, 1975). With lexical expansion, production accuracy drops but eventually their accuracy rises again. This is the so-called u-shaped learning curve (Ferguson & Farwell, 1975). Faes and Gillis (2020) found that ABI1's first word productions are indeed more accurate than later acquired words. However, inspection of the rating scale and transcription task results seemed to indicate that this did not result in intelligible speech for unfamiliar listeners. For children with CI and TH, the early words were indeed judged to be the most intelligible, and therefore presumably also the most accurately produced. So, the

children with CI and TH seem to follow the suggested u-shaped learning postulated by Ferguson and Farwell (1975). As lexicon size increased, the difference between the children with ABI and the control groups enlarged. Speech intelligibility performance of ABI1 and ABI3 was well below that of children with CI and TH with lexical expansion. But, in contrast, speech intelligibility performance of ABI2 approached that of children with CI and TH, especially with increasing lexicon size to more than 500 word types.

Moderate levels of speech intelligibility in all groups

With lexical development, speech intelligibility slightly increased for the children with CI, the children with TH and ABI2, though their performance was still fairly modest. For ABI1 and ABI3, however, little to no change was observed with lexical expansion. So, overall, also for children with CI and TH, speech intelligibility seemed only moderate. Different factors may account for this result. A first important factor is the expertise of the listener or transcriber with child language in general. For instance Munson et al. (2012) showed that listeners benefit from the experience with specific types of speech, and Hustad and Cahill (2003) showed increased intelligibility of dysarthric speech when listeners were familiarized with it through different trials. The listeners in our study were unfamiliar with the speech of children with hearing problems. Yet, the familiarity with child speech in general did not affect the outcomes in the present study (see method section). Similarly, also for instance Boonen et al. (2019) did not find an effect of listener's background (e.g. speech and language therapist as compared to primary school teachers and inexperienced listeners) on their ability to identify children with CI and children with TH.

Secondly, listeners were unfamiliar to the specific children in the present study. Since each child's speech has its own characteristics, the task is assumed to be more difficult for unfamiliar listeners than for listeners familiar with the child (Cox & McDaniel, 1989).

Therefore, the overall intelligibility judgements may be also fairly modest in children with TH and CI in this experiment. For children with ABI, it has been shown that the best performing children reach intelligible speech only for familiar, experienced listeners (Aslan et al., 2020; Sennaroglu, Sennaroglu, et al., 2016; van der Straaten et al., 2019). So, listening to the children with ABI must have been very challenging for our (inexperienced) listeners unfamiliar with the children in this study.

Thirdly, the stimuli in this experiment have been presented to the listeners without any kind of contextual information, which complicates the task tremendously. Different contextual facets were blinded to the listeners. For instance, the listeners had only access to audio files. Studies showed that speech intelligibility of different clinical groups of speakers increased when stimuli were presented in an audiovisual mode rather than an audio-only mode (Hubbard & Kushner, 1980; Keintz et al., 2007). Moreover, speech intelligibility in this study was judged based on single-word productions. Research suggests that more contextualized utterances (e.g. sentences or short conversations) are judged to be more intelligible than utterances with less contextual information, such as the one-word utterances in the present study (Baudonck et al., 2010; Boonen, 2020; Montag et al., 2014). In sentences, other understood words may hint to the general subject of the sentence, which makes the utterance more intelligible overall (Hustad et al., 2020). In conversations, listeners may find support in the interaction with adults and other children, improving the intelligibility judgements as well. In this experiment, the condition was isolated words, with no aid of any context whatsoever, which may also explain the moderate intelligibility scores in the entire experiment. Yet, Baudonck et al. (2010) advocated that testing intelligibility at the word level is more sensitive to subtle differences in children's speech intelligible, precisely because of the lack of contextual information. It can thus be considered a stricter measure of speech intelligibility.

Finally, the stimuli of children with ABI in this study came from children using their device between one and four years, with chronological ages between three and six years. A typically developing child (with TH) is only entirely intelligible by four years of age (Hustad et al., 2020). For children with CI, it takes even longer (Chin & Kuhns, 2014), even though in SIR's terminology, they could reach ceiling scores earlier on (De Raeve, 2010). For the better performing children with ABI, ceiling scores have not been observed even after five to six years of device use (Aslan et al., 2020; Sennaroglu, Sennaroglu, et al., 2016; van der Straaten et al., 2019). Rather, they can reach a score of 3 or 4 on the SIR (i.e. being intelligible to familiar listeners with or without lip-reading) by that time. So, listening to children with ABI with only one to four years of device use in this study has inevitably been very challenging for the untrained, unfamiliar listeners, since not even children with TH are completely intelligible at these early hearing ages.

Implications

Good speech intelligibility skills have been repeatedly related to better psychosocial functioning of children with TH and CI (V. Freeman, D. Pisoni, et al., 2017; Most et al., 2012; Preisler et al., 2002; Wong et al., 2016). Even though not studied yet, it seems reasonable that speech intelligibility and psychosocial well-being are linked in the ABI group as well. The early intelligibility results of our children with ABI (and the differences between the children) can be highly informative for their future development. In another clinical population, i.e. children with cerebral palsy, Hustad et al. (2019) showed that the early speech intelligibility scores at three years of age predicted the speech intelligibility outcomes at eight years of age. The chronological age of the children with ABI in this study was three to four years of age for the early data points (up to 100 word types) and five to six years of age from ca. 350 word types onwards (see Table 1). Since there was little to no development in the

children with ABI's speech intelligibility, except for ABI2, this may not be a good indication for these children's future development and, consequently, also their psychosocial well-being. In addition, the literature showed that a child with TH is intelligible for an unfamiliar listeners by four years of age (Hustad et al., 2020). The children with ABI in this study did not reach this level, even when they were two years older at the end of the study.

Even though it seems, e.g. for ABI2, that the difference with children with CI and children with TH is quite acceptable, this must be seen in the light of the measure of comparison used. All groups were matched on their lexical age, for reasons such as the close link between phonology and lexicon (e.g. Stoel-Gammon, 2011) and intelligibility and phonology (Ingram, 2002). However, as a result, a two-year old child with TH was matched to, for instance, five-year old children with ABI (see Table 1). The same holds for hearing age and children with CI. Children with CI are implanted at least one year earlier than the children with ABI and thus have at least one year of device use more than the children with ABI at similar chronological ages. These age and hearing age differences skew the comparisons considerably. As stated in the introduction, there are different options to match the groups, but none of the options will ever be optimal. Importantly, it should be kept in mind that even though ABI2 scores somewhat similar to the children with TH and CI, this child is considerably older at that time. By the end of the data collection, this six-year-old child with ABI was performing more or less similar to a two-year old child with TH and a three to four-year-old child with CI. So, in terms of chronological age, this child is still lagging behind age-mates with TH and CI. A similar picture holds for ABI1 and ABI3, of course, who are lagging behind even more. This is also confirmed by studies using hearing age as a measure of comparison. For instance in Faes and Gillis (2020), it is shown that the same children with ABI produced their words significantly less accurately than TH and CI peers matched on

hearing age. Hearing ages of the children with TH and CI in this study were at least one year lower in the group matching, which may thus heavily impact the results.

Interindividual variation

There was a considerable difference in the judged intelligibility of the three children with ABI, with ABI2 outperforming the other two children. One aspect that may have contributed to this difference is the children's length of device use and chronological age differences between the children with ABI. At ca. 350 word types, for instance, ABI3 used his ABI device one year less than the other two children (with equal length of device use though, see Table 1). This may have caused lower intelligibility scores for ABI3 as compared to ABI2, but does not explain the difference between ABI1 and ABI2. Still, ABI3 had a similar lexicon size than ABI1 and ABI2 with this shorter ABI hearing time, which may have resulted from the child's CI use before ABI implantation. So, in terms of lexical development, ABI3 may be benefitting from the period with CI use, and possibly also from the combination of the ABI and the CI as suggested by Friedman et al. (2018) and Batuk et al. (2020). Nevertheless, this did not result in spoken intelligibility performance similar to that of children with CI and TH.

ABI2 outperformed the other children with ABI in speech intelligibility performance, but also in lexicon expansion. For instance, after two years of hearing age, ABI1 had ca. 100 word types, and ABI2 ca. 200 word types. Thus far, it is unclear which factors contribute to these individual differences. ABI1 and ABI2, for instance, were implanted at approximately the same chronological age, had an equal number of activated electrodes, and relatively similar hearing thresholds after implantation. As in other clinical populations, such as children with CI (e.g. Duchesne et al., 2009; Svirsky et al., 2000; Szagun, 2002; Wie, 2010), it may be the case that the interindividual variation is larger in the ABI population than in typically developing children, in which individual variation is also present though (e.g. Hustad et al.,

2020 for speech intelligibility). In this respect, Nagels et al. (2020) advocated for the investigation of individual patterns in “heterogeneous clinical populations such as CI users” (p. 286) to improve speech and language therapy. Similarly, also Pisoni et al. (2017) highlighted the importance of individual variation in CI research. Our results seem to confirm the importance of individual analyses in the ABI population as well, given the differences found between the three children.

Concluding remarks

In this study, scarce data of individual cases of a specific subpopulation (the three children with ABI and their utterances) were used and a larger sample of control subjects assumed to represent a population of control subjects (utterances of children with CI and children with TH at particular points in their lexical development). The statistical analyses took advantage of the richer dataset of control subjects to estimate population averages and accompanying variance estimates, combined with the data of the three cases of children with ABI. This enabled modeling how the individual cases were positioned in comparison with population estimates. The analyses were not presented as estimates of population level characteristics for all children with ABI. Rather the individual ABI estimates were presented as explorative case-level data, giving the analyses a more explorative character, but still making use of the statistical power of the control group data to put individual data into perspective. This method deemed fruitful in our study and we suggest that scholars who also study very specific phenomena for which it is difficult to gather broad sample data but rather unique case data to consider this method if valuable sample data are available to use as a point of comparison.

To conclude, our results suggest that speech intelligibility of children with ABI is susceptible to considerable individual variation. Whereas one child with ABI seemed to approach the CI and TH levels of speech intelligibility, the other two children with ABI

remained well below the control groups matched on lexicon size. Overall, speech intelligibility was only moderate in all groups of children, with quite low rating scores on the 100-point scale and large differences in the listeners' transcriptions.

References

- AlSanosi, A., & Hassan, S. M. (2014). The effect of age at cochlear implantation outcomes in Saudi children. *International Journal of Pediatric Otorhinolaryngology*, 78(2), 272 - 276. <https://doi.org/10.1016/j.ijporl.2013.11.021>
- Archbold, S., Lutman, M. E., & Marshall, D. H. (1995). Categories of Auditory Performance. *Annals of Otology, Rhinology & Laryngology*, 166 Suppl, 312 - 314.
- Aslan, F., Burcu Ozkan, H., Yücel, E., Sennaroglu, G., Bilginer, B., & Sennaroglu, L. (2020). Effects of age at auditory brainstem implantation: impact on auditory perception, language development, speech intelligibility. *Otology & Neurotology*, 41(1), 11 - 20. <https://doi.org/10.1097/MAO.0000000000002455>
- Baayen, H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge University Press.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1 - 48. <https://doi.org/10.18637/jss.v067.i01>
- Batuk, O. M., Cinar, B. C., Yarali, M., Aslan, F., Ozkan, H. B., Sennaroglu, G., Yucel, E., Bajin, M. D., Bilginer, B., & Sennaroglu, L. (2020). Bimodal stimulation in children with inner ear malformation: One side cochlear implant and contralateral auditory brainstem implant. *Clinical Otolaryngology*, 45, 231 - 238. <https://doi.org/10.1111/coa.13499>

- Baudonck, N., Buekers, R., Gillebert, S., & Van Lierde, K. (2009). Speech intelligibility of Flemish children as judged by their parents. *Folia Phoniatrica et Logopaedica*, 61(5), 288 - 295. <https://doi.org/10.1159/000235994>
- Baudonck, N., Dhooge, I., & Van Lierde, K. (2010). Intelligibility of hearing impaired children as judged by their parents: a comparison between children using cochlear implants and children using hearing aids. *International Journal of Pediatric Otorhinolaryngology*, 74, 1310-1315. <https://doi.org/10.1016/j.ijporl.2010.08.011>
- Bayazit, Y., Kosaner, J., Cicek Cinar, B., Atac, A., Tutar, H., Gunduz, B., Altinyay, S., Gokdogan, C., Ant, A., Ozdek, A., & Goksu, N. (2014). Methods and preliminary outcomes of pediatric auditory brainstem implantation. *Annals of Otolaryngology, Rhinology & Laryngology*, 123(8), 529 - 536. <https://doi.org/10.1177/0003489414525123>
- Boonen, N. (2020). *Identifiability and intelligibility of the speech of children with a cochlear implant : a comparison with normally hearing children and children with an acoustic hearing aid* [University of Antwerp]. Antwerp.
- Boonen, N., Kloots, H., Verhoeven, J., & Gillis, S. (2019). Can listeners hear the difference between children with normal hearing and children with a hearing impairment? *Clinical Linguistics & Phonetics*, 33, 316 - 333. <https://doi.org/10.1080/02699206.2018.1513564>
- Brown, R. (1973). *A first language. The early stages*. George Allen & Unwin Ltd.
- Calmels, M., Saliba, I., Wanna, G., Cochard, N., Fillauw, J., Deguine, O., & Fraysse, B. (2004). Speech perception and speech intelligibility in children after cochlear implantation. *International Journal of Pediatric Otorhinolaryngology*, 68, 347 - 351. <https://doi.org/10.1016/j.ijporl.2003.11.006>
- Caselli, M., Rinaldi, P., Varuzza, C., Giuliani, A., & Burdo, S. (2012). Cochlear implant in the second year of life: lexical and grammatical outcomes. *Journal of Speech, Language and Hearing Research*, 55, 382 - 394. [https://doi.org/10.1044/1092-4388\(2011/10-0248\)](https://doi.org/10.1044/1092-4388(2011/10-0248))

- Chin, S., Bergeson, T., & Phan, J. (2012). Speech intelligibility and prosody production in children with cochlear implants. *Journal of Communication Disorders*, 45, 355-366. <https://doi.org/10.1016/j.jcomdis.2012.05.003>
- Chin, S., & Kuhns, M. (2014). Proximate factors associated with speech intelligibility in children with cochlear implants: A preliminary study. *Clinical Linguistics & Phonetics*, 28, 532 - 542. <https://doi.org/10.3109/02699206.2014.926997>
- Chin, S., Tsai, P., & Gao, S. (2003). Connected speech intelligibility of children with cochlear implants and children with normal hearing. *American Journal of Speech-Language Pathology*, 12, 440-451. [https://doi.org/10.1044/1058-0360\(2003/090\)](https://doi.org/10.1044/1058-0360(2003/090))
- Chomsky, N. (1957). *Syntactic structures*. Mouton.
- Colletti, L., Shannon, R., & Colletti, V. (2014). The development of auditory perception in children following auditory brainstem implantation. *Audiology and Neurotology*, 19(6), 386 - 394. <https://doi.org/10.1159/000363684>
- Colletti, V., Fiorino, F., Sacchetto, L., Miorelli, V., & Carner, M. (2001). Hearing habilitation with auditory brainstem implantation in two children with cochlear nerve aplasia. *International Journal of Pediatric Otorhinolaryngology*, 60, 99 - 111. [https://doi.org/10.1016/S0165-5876\(01\)00465-7](https://doi.org/10.1016/S0165-5876(01)00465-7)
- Cox, R. M., & McDaniel, D. M. (1989). Development of the Speech Intelligibility Rating (SIR) test for hearing aid comparisons. *Journal of Speech and Hearing Research*, 32, 347 - 352. <https://doi.org/10.1044/jshr.3202.347>
- De Raeve, L. (2010). A longitudinal study on auditory perception and speech intelligibility in deaf children implanted younger than 18 months in comparison to those implanted at later ages. *Otology & Neurotology*, 31(8), 1261 - 1267. <https://doi.org/10.1097/MAO.0b013e3181f1cde3>

- Duchesne, L., Sutton, A., & Bergeron, F. (2009). Language achievement in children who received cochlear implants between 1 and 2 years of age: group trend and individual patterns. *Journal of Deaf Studies and Deaf Education*, 14(4), 465 - 485. <https://doi.org/10.1093/deafed/enp010>
- Eisenberg, L., Hammes Gangly, D., Martinez, A., Fisher, J. M., Winter, M., Glater, J., Schrader, D. K., Loggins, J., Wilkinson, E., & The Los Angeles Pediatric ABI Team. (2018). Early communication development of children with auditory brainstem implants. *Journal of Deaf Studies and Deaf Education*, 23(3), 249 - 260. <https://doi.org/10.1093/deafed/eny010>
- Ertmer, D., & Goffman, L. (2011). Speech production accuracy and variability in young cochlear implant recipients: comparisons with typically developing age-peers. *Journal of Speech, Language, and Hearing Research*, 54, 177 - 189. [https://doi.org/10.1044/1092-4388\(2010/09-0165\)](https://doi.org/10.1044/1092-4388(2010/09-0165))
- Faes, J., Boonen, N., & Gillis, S. (2019). Hersenstamimplantatie bij kinderen: wat is het effect op de vroege spraakproductie? *Logopedie*, 32, 9 - 19.
- Faes, J., & Gillis, S. (2016). Word initial fricative production in children with cochlear implants and their normally hearing peers matched on lexicon size. *Clinical Linguistics & Phonetics*, 30(12), 959 - 982. <https://doi.org/10.1080/02699206.2016.1213882>
- Faes, J., & Gillis, S. (2019a). Auditory brainstem implantation in children: effect on speech production. *International Journal of Pediatric Otorhinolaryngology*, 119, 103 - 112. <https://doi.org/10.1016/j.ijporl.2019.01.014>
- Faes, J., & Gillis, S. (2019b). Expressive vocabulary growth after pediatric auditory brainstem implantation in two cases' spontaneous productions: a comparison with children with cochlear implants and typical hearing. *Frontiers in Pediatrics*. <https://doi.org/10.3389/fped.2019.00191>

- Faes, J., & Gillis, S. (2020). Word characteristics and speech production accuracy in children with auditory brainstem implants: a longitudinal triple case report. *Clinical Linguistics & Phonetics, OnlineFirst*. <https://doi.org/10.1080/02699206.2020.1838613>
- Faes, J., & Gillis, S. (2021). Consonant and vowel production in the spontaneous speech productions of children with auditory brainstem implants. *Clinical Linguistics & Phonetics, In press*. <https://doi.org/10.1080/02699206.2020.1869833>
- Ferguson, C., & Farwell, C. (1975). Words and sounds in early language development. *Language, 51*(2), 419 - 439.
- Fernandes, N. F., de Queiroz Teles Gomes, M., Tsuji, R., Bento, R., & Goffi-Gomez, M. (2020). Auditory and language skills in children with auditory brainstem implants. *International Journal of Pediatric Otorhinolaryngology, 132*, 110010. <https://doi.org/10.1016/j.ijporl.2020.110010>
- Flipsen, P., & Colvard, L. (2006). Intelligibility of conversational speech produced by children with cochlear implants. *Journal of Communication Disorders, 39*, 93 - 108.
- Freeman, S., Pisoni, D., Kronenberger, W., & Castellanos, I. (2017). Speech intelligibility and psychosocial functioning in deaf children and teens with cochlear implants. *Journal of Deaf Studies and Deaf Education, 22*(3), 278 - 289. <https://doi.org/10.1093/deafed/enx001>
- Freeman, V., Pisoni, D., Kronenberger, W., & Castellanos, I. (2017). Speech intelligibility and psychosocial functioning in deaf children and teens with cochlear implants. *Journal of Deaf Studies and Deaf Education, 22*(3), 278 - 289. <https://doi.org/10.1093/deafed/enx001>
- Freeman, V., Pisoni, D. B., Kronenberger, W. G., & Castellanos, I. (2017, Jul 01). Speech intelligibility and psychosocial functioning in deaf children and teens with cochlear

- implants. *Journal of Deaf Studies and Deaf Education*, 22(3), 278-289.
<https://doi.org/https://doi.org/10.1093/deafed/enx001>
- Friedman, D. R., Asfour, L., Shapiro, W., Thomas Roland, J., & Waltzman, S. (2018). Performance with an auditory brainstem implant and contralateral cochlear implant in pediatric patients. *Audiology and Neurotology*, 23, 216 - 221.
<https://doi.org/10.1159/000493085>
- Frinsel, F., Kingma, A., Swarte, F., & Gooskens, C. (2015). Predicting the asymmetric intelligibility between spoken Danish and Swedish using conditional entropy. *Tijdschrift voor Skandinavistiek*, 34(2), 120 - 138.
- Gordon-Brannan, M., & Hodson, B. W. (2000). Intelligibility/severity measurements of prekindergarten children's speech. *American Journal of Speech-Language Pathology*, 9(2). <https://doi.org/10.1044/1058-0360.0902.141>
- Grandon, B., Martinez, M., Samson, A., & Vilain, A. (2020). Long-term effects of cochlear implantation on the intelligibility of speech in French-speaking children. *Journal of Child Language*, 47(4), 881 - 892. <https://doi.org/10.1017/S0305000919000837>
- Habib, M. G., Waltzman, S., Tajudeen, B., & Svirsky, M. (2010). Speech production intelligibility of early implanted pediatric cochlear implant users. *International Journal of Pediatric Otorhinolaryngology*, 74(8), 855 - 859.
<https://doi.org/10.1016/j.ijporl.2010.04.009>
- Hubbard, D., & Kushner, D. (1980). A comparison of speech intelligibility between esophageal and normal speakers via three modes of presentation. *Journal of Speech, Language and Hearing Research*, 23(4), 909 - 916. <https://doi.org/10.1044/jshr.2304.909>
- Hustad, K. C., & Cahill, M. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology*, 12(2), 198 - 208. [https://doi.org/10.1044/1058-0360\(2003/066\)](https://doi.org/10.1044/1058-0360(2003/066))

- Hustad, K. C., Mahr, T. J., Broman, A. T., & Rathouz, P. J. (2020). Development of speech intelligibility between 30 and 47 months in typically developing children: A cross-sectional study of growth. *Journal of Speech Language and Hearing Research*, 63(1675 - 1687), 32 - 48. https://doi.org/10.1044/2020_JSLHR-20-00008
- Hustad, K. C., Sakash, A., Natzke, P., Broman, A. T., & Rathouz, P. J. (2019). Longitudinal growth in single word intelligibility among children with cerebral palsy from 24 to 96 months of age: Predicting later outcomes from early speech production. *Journal of Speech, Language and Hearing Research*, 62, 1599 - 1613. https://doi.org/10.1044/2018_JSLHR-S-18-0319
- Ingram, D. (2002). The measurement of whole-word productions. *Journal of Child Language*, 29, 713-733. <https://doi.org/10.1017/S0305000902005275>
- Kaplan, A., Kozin, E., Puram, S., Owoc, M., Shah, P., Hight, A., Sethi, R., Remenschneider, A., & Lee, D. (2015). Auditory brainstem implant candidacy in the United States in children 0-17 years old. *International Journal of Pediatric Otorhinolaryngology*, 79, 310 - 315. <https://doi.org/10.1016/1.ijporl.2014.11.023>
- Keintz, C., Bunton, K., & Hoit, J. (2007). Influence of visual information on the intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology*, 16(3), 222 - 234. [https://doi.org/10.1044/1058-0360\(2007/027\)](https://doi.org/10.1044/1058-0360(2007/027))
- Kidd, E., & Donnelly, S. (2020). Individual differences in first language acquisition. *Annual Review of Linguistics*, 6, 319 - 340.
- Knowles, J., & Frederick, C. (2020). merTools: Tools for Analyzing Mixed Effect Regression Models. [online via:] <https://CRAN.R-project.org/package=merTools>.
- Lagerberg, T., Asberg, J., Hartelius, L., & Persson, C. (2014). Assessment of intelligibility using children's spontaneous speech: Methodological aspects. *International Journal of*

- Language & Communication Disorders*, 49(2), 228 - 239. <https://doi.org/10.1111/1460-6984.12067>
- Lejeune, B., & Demanez, L. (2006). Speech discrimination and intelligibility: Outcome of deaf children fitted with hearing aids or cochlear implants. *B-ENT*, 2(2), 63 - 68.
- MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk*. NJ: Lawrence Erlbaum Associates.
- Miller, N. (2013). Measuring up to speech intelligibility. *International Journal of Language & Communication Disorders*, 48(6), 601 - 612. <https://doi.org/10.1111/1460-6984.12061>
- Moberg, J., Gooskens, C., Nerbonne, J., & Vaillette, N. (2007). Conditional entropy measures intelligibility among related languages. In P. Dirix, I. Schuurman, V. Vandeghinste, & F. Van Eynde (Eds.), *Computational Linguistics in the Netherlands 2006: Selected papers from the 17th CLIN Meeting* (pp. 51 - 66). LOT.
- Montag, J., AuBuchon, A., Pisoni, D., & Kronenberger, W. (2014). Speech intelligibility in deaf children after long-term cochlear implant use. *Journal of Speech, Language and Hearing Research*, 57, 2332 - 2343. https://doi.org/10.1044/2014_JSLHR-H-14-0190
- Most, T., Ingber, S., & Heled-Ariam, E. (2012). Social competence, sense of loneliness, and speech intelligibility of young children with hearing loss in individual inclusion and group inclusion. *Journal of Deaf Studies and Deaf Education*, 17(2), 259 - 272. <https://doi.org/10.1093/deafed/enr049>
- Munson, B., Johnson, J. M., & Edwards, J. (2012). The role of experience in the perception of phonetic detail in children's speech: a comparison between speech-language pathologists and clinically untrained listeners. *American Journal of Speech-Language Pathology*, 21(2), 124 - 139. [https://doi.org/10.1044/1058-0360\(2011/11-0009\)](https://doi.org/10.1044/1058-0360(2011/11-0009))

- Nagels, L., Bastiaanse, R., Baskent, D., & Wagner, R. A. (2020). Individual differences in lexical access among cochlear implant users. *Journal of Speech Language and Hearing Research*, 63, 286 - 304. https://doi.org/10.1044/2019_JSLHR-19-00192
- Nicholson, H., Munson, B., Reidy, P., & Edwards, J. (2015, 08/2015). *Effects of age and vocabulary size on production accuracy and acoustic differentiation of young children's sibilant fricatives* International Congress of Phonetic Sciences, Glasgow, Scotland.
- Peng, S., Spencer, L. J., & Tomblin, B. (2004). Speech intelligibility of pediatric cochlear implant recipients with 7 years of device experience. *Journal of Speech, Language and Hearing Research*, 47(6), 1227 - 1236. [https://doi.org/10.1044/1092-4388\(2004\)092](https://doi.org/10.1044/1092-4388(2004)092)
- Pisoni, D., Kronenberger, W., Harris, M., & Moberly, A. (2017). Three challenges for future research on cochlear implants. *World Journal of Otorhinolaryngology - Head and Neck Surgery*, 3, 240 - 254. <https://doi.org/10.1016/j.wjorl.2017.12.010>
- Preisler, G., Tvingstedt, A., & Ahlström, M. (2002). A psychosocial follow-up study of deaf preschool children using cochlear implants. *Child: Care, Health & Development*, 28(5), 403 - 418. <https://doi.org/10.1046/j.1365-2214.2002.00291.x>
- Puram, S., & Lee, D. (2015). Pediatric auditory brainstem implant surgery. *Otolaryngologic Clinics of North America*, 48, 1117 - 1148. <https://doi.org/10.1016/j.otc.2015.07.013>
- R Core Team. (2013). *R: a language and environment for statistical computing*. In R foundation for statistical computing. <http://www.R-project.org/>
- Reidy, P., Beckman, M., Litovsky, R., & Edwards, J. (2015). The acquisition of English sibilant fricatives by children with bilateral cochlear implants. International congress of phonetic sciences, Glasgow.
- Santos, B., & Sosa, A. (2015). *Measuring the relationship between lexical and phonological development in typically developing 2- and 3-year-olds* Poster presented at the International child phonology conference 2015, Canada, 24 - 26 June 2015.

- Schramm, B., Bohnert, A., & Keilmann, A. (2010). Auditory, speech and language development in young children with cochlear implants compared with children with normal hearing. *International Journal of Pediatric Otorhinolaryngology*, 74, 812 - 819. <https://doi.org/10.1016/j.ijporl.2010.04.008>
- Sennaroglu, L., Colletti, V., Lenarz, T., Manrique, M., Laszig, R., Rask-Andersen, H., Gösku, N., Offeciers, E., Saeed, S., Behr, R., Bayazit, Y., Casselman, J., Freeman, S., Kileny, P., Lee, D., Shannon, R., Kameswaran, M., Hagr, A., Zarowski, A., Schwartz, M., Bilginer, B., Kishore, A., Sennaroglu, G., Yücel, E., Sarac, S., Atas, A., Colletti, L., O'Driscoll, M., Moon, I., Gärtner, L., Huarte, A., Nyberg, G., Özgen Mocan, B., Atay, G., Demir Bajin, M., Cicek Cinar, B., Özbal Batuk, M., Yarali, M., Aydinli, F., Aslan, F., Kirazli, M., Özkan, B., Hans, J., Kosaner, J., & Polak, M. (2016). Consensus statement: long-term results of ABI in children with complex inner ear malformations and decision making between CI and ABI. *Cochlear Implants International*, 17(4), 163 - 171. <https://doi.org/10.1080/14670100.2016.1208396>
- Sennaroglu, L., Sennaroglu, G., Yücel, E., Bilginer, B., Atay, G., Demir Bajin, M., Özgen Mocan, B., Yarali, M., Aslan, F., Cicek Cinar, B., Özkan, B., Özbal Batuk, M., Ekin Kirazli, C., Karakaya, J., Atas, A., Sarac, S., & Ziyal, I. (2016). Long-term results of ABI in children with severe inner ear malformations. *Otology & Neurotology*, 37, 865 - 872. <https://doi.org/10.1097/MAO.0000000000001050>
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379 - 423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Sosa, A., & Stoel-Gammon, C. (2006). Patterns of intra-word phonological variability during the second year of life. *Journal of Child Language*, 33, 31 - 50. <https://doi.org/10.1017/S0305000905007166>

- Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, 38, 1 - 34. <https://doi.org/10.1017/S0305000910000425>
- Strömbergsson, S., Holm, K., Edlund, J., Lagerberg, T., & McAllister, A. (2020). Audience Response System-Based Evaluation of Intelligibility of Children's Connected Speech-Validity, Reliability and Listener Differences. *Journal of Communication Disorders*, 87, 106037. <https://doi.org/10.1016/j.jcomdis.2020.106037>
- Svirsky, M., Robbins, A., Kirk, K., Pisoni, D., & Miyamoto, R. (2000). Language development in profoundly deaf children with cochlear implants. *Psychological Science*, 11(2), 153 - 157. <https://doi.org/10.1111/1467-9280.00231>
- Szagun, G. (2002). The acquisition of grammar in young German-speaking children with cochlear implants and with normal hearing. *Antwerp papers in linguistics*, 102, 40 - 60.
- Teagle, H. F. B., Henderson, L., He, S., Ewend, M. G., & Buchman, C. A. (2018). Pediatric auditory brainstem implantation: surgical, electrophysiologic, and behavioral outcomes. *Ear & Hearing*, 39, 326 - 336. <https://doi.org/10.1097/AUD.0000000000000501>
- Toe, D., & Paatsch, L. (2013). The conversational skills of school-aged children with cochlear implants. *Cochlear Implants International*, 14(2), 67 - 79. <https://doi.org/10.1179/1754762812Y.0000000002>
- Tseng, S. C., Kuei, K., & Tsou, P. C. (2011). Acoustic characteristics of vowels and plosives/affricates of Mandarin-speaking hearing-impaired children. *Clinical Linguistics & Phonetics*, 25(9), 784 - 803. <https://doi.org/10.3109/02699206.2011.565906>
- 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* [PhD, University of Antwerp]. Antwerp.

- van der Straaten, T., Netten, A., Boermans, P., Briaire, J., Scholing, E., Koot, R., Malessy, M., van der Mey, A., Verbist, B., & Frijns, J. (2019). Pediatric auditory brainstem implant users compared with cochlear implant users with additional disabilities. *Otology & Neurotology*, *40*(7), 936 - 945. <https://doi.org/10.1097/MAO.0000000000002306>
- van Heuven, V. J. (2008). Making sense of strange sounds: (Mutual) intelligibility of related language varieties. A review. *International Journal of Humanities and Arts Computing*, *2*(1-2), 39-62. <https://doi.org/https://doi.org/10.3366/e1753854809000305>
- Whitehill, T. L., & Ciocca, V. (2000, Dec). Perceptual-phonetic predictors of single-word intelligibility: A study of Cantonese dysarthria. *Journal of Speech, Language, and Hearing Research*, *43*(6), 1451-1465. <https://doi.org/https://doi.org/10.1044/jslhr.4306.1451>
- Wie, O. (2010). Language development in children after receiving bilateral cochlear implants between 5 and 18 months. *International Journal of Pediatric Otorhinolaryngology*, *74*, 1258 - 1266. <https://doi.org/10.1016/j.ijporl.2010.07.026>
- Wilkinson, E., Eisenberg, L., Krieger, M., Schwartz, M., Winter, M., Glater, J., Martinez, A., Fishher, L., Shannon, R., & The Los Angeles Pediatric ABI Team. (2017). Initial results of a safety and feasibility study of audiotry brainstem implantation in congenitally deaf children. *Otology & Neurotology*, *38*(2), 212 - 220. <https://doi.org/10.1097/MAO.0000000000001287>
- Wong, C., Ching, T., Leigh, G., Cupples, L., Button, L., Marnane, V., Whitfield, J., Gunnourie, M., & Martin, L. (2016). Psychosocial development of 5-year-old children with hearing loss: risks and protective factors. *International Journal of Audiology, Early Online*, 1 - 12. <https://doi.org/10.1080/14992027.2016.1211764>

- Wong, K., Kozin, E., Kanumuri, V., Vachicouras, N., Miller, J., Lacour, S., Brown, C., & Lee, D. (2019). Auditory brainstem implants: recent progress and future perspectives. *Frontiers in Neuroscience, 13:10*. <https://doi.org/10.3389/fnins.2019.00010>
- Yucel, E., Aslan, F., Ozkan, H., & Sennaroglu, L. (2015). Recent rehabilitation experience with pediatric ABI users. *The Journal of International Advanced Otology, 11*, 110 - 113. <https://doi.org/10.5152/iao.2015.915>

Tables

Table 1. Grouping of the ABI data according to the different measures of comparison.

Grouping point	Child ID	Data point (Number per child)	Chronological age (years;months)	Hearing age with ABI (years;months)	Cumulative vocabulary size (word types)
1. < 50 word types	ABI1	1	3;05	1;05	16
2. ca. 100 word types	ABI1	2	4;04	2;03	78
	ABI3	1	4;03	0;03	90
3. ca. 200 word types	ABI2	1	4;04	2;03	183
4. ca. 350 word types	ABI1	3	5;03	3;03	324
	ABI2	2	5;03	3;02	394
	ABI3	2	5;04	1;04	315
5. > 500 word types	ABI2	3	6;03	4;02	611

Table 2. Individual data of the children with CI

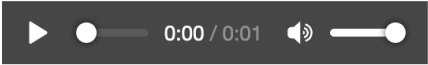
ID	Gender	PTA unaided (dB HL)	PTA CI (dB HL) (age 2;00)	Age CI implantation	Age second CI
CI1	F	120	48	1;01	6;03
CI2	F	120	30	0;07	4;08
CI3	F	115	33	0;10	5;10
CI4	M	113	48	1;06	-
CI5	M	93	38	1;05	6;04
CI6	M	120	53	0;09	-
CI7	F	117	42	0;05	1;03
CI8	F	112	38	1;07	-
CI9	F	103	28	0;08	1;11

dB HL = decibel Hearing Level

Ages are presented in years;months

- = no second CI

Figures.



How intelligible is the utterance according to you?



What do you think the child is saying?

Figure 1. Example of the listeners' task in Qualtrics

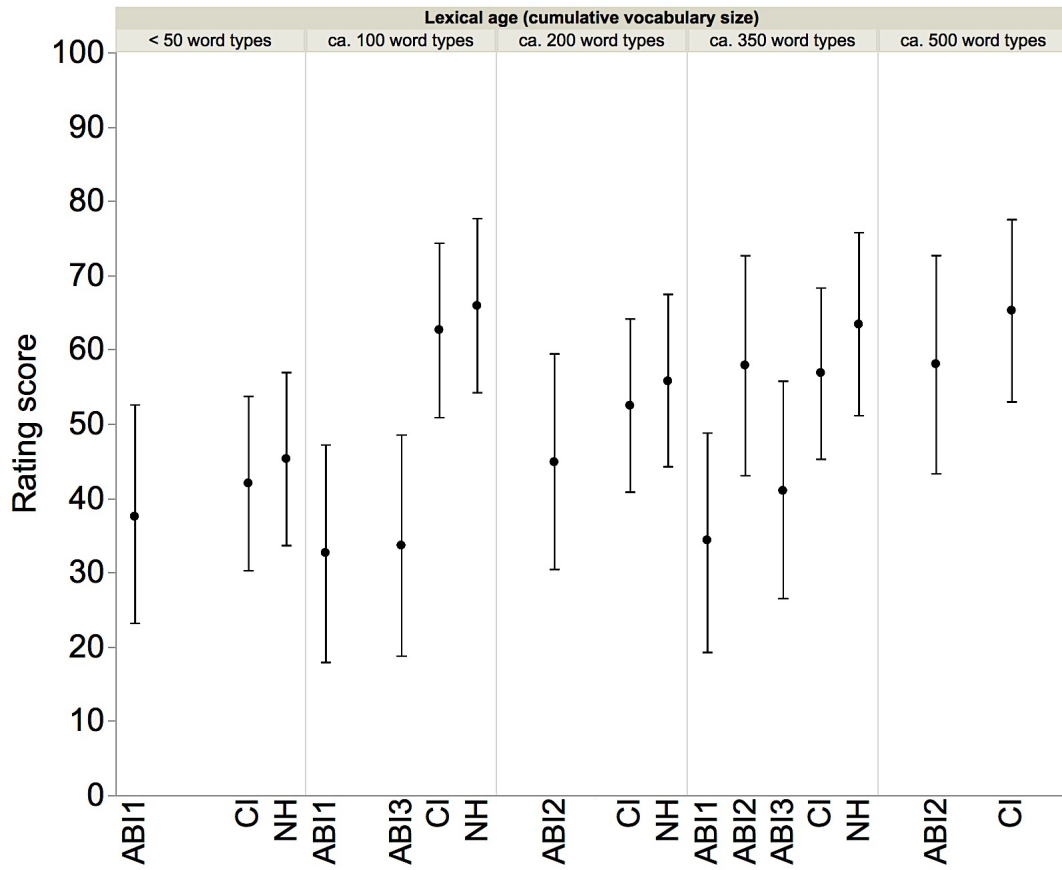


Figure 2. Intelligibility rating score relative to lexical age – predicted intervals

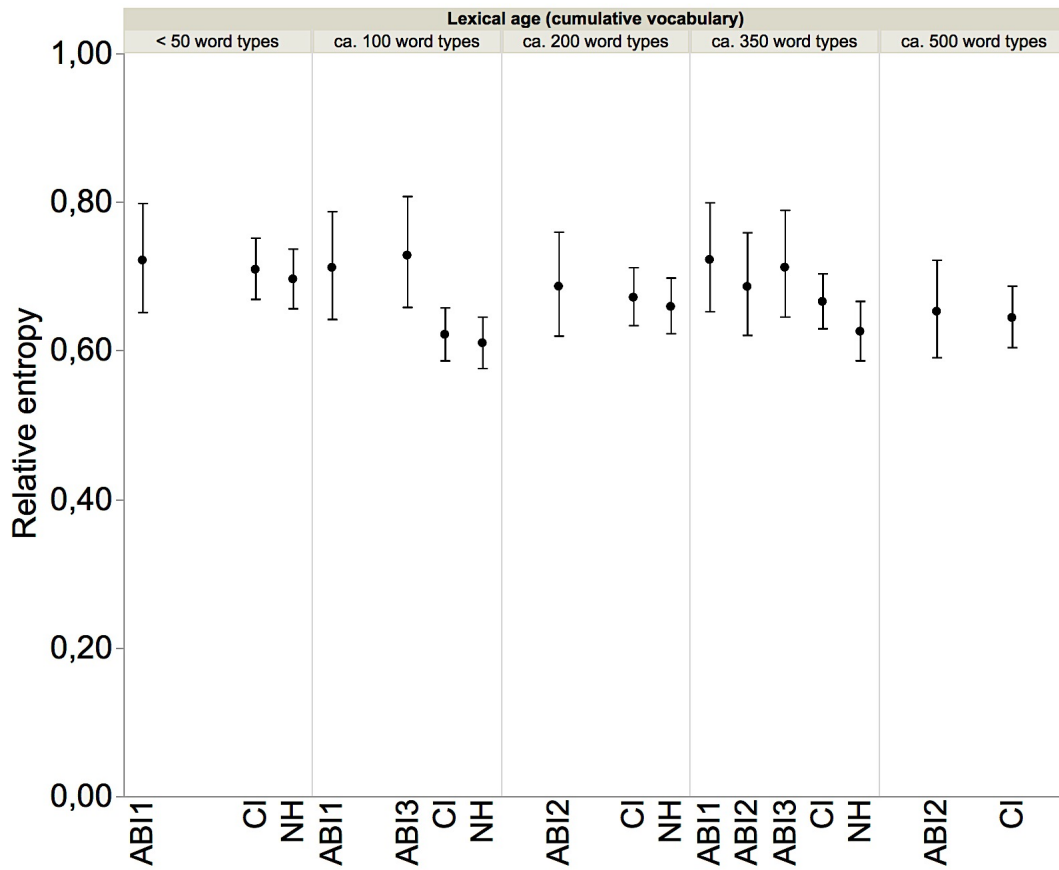


Figure 3. Relative entropy (transcription task) relative to lexical age – predicted intervals

Figure captions.

Figure 1. Example of the listeners' task in Qualtrics

Figure 2. Intelligibility rating score relative to lexical age – predicted intervals

Figure 3. Relative entropy (transcription task) relative to lexical age – predicted intervals

Supplementary materials.

Table S1. Precise matching of children with ABI with children with CI.

	Cumulative vocabulary size ABI group	Number of CI files matched	Child IDs	Mean cumulative vocabulary size (SD)	Mean chronological age (SD) (years;months)
1.	< 50 word types	3	CI4_1 CI5_1 CI8_1	10.75 (10.31)	1;11 (0;01)
2.	ca. 100 word types	3	CI4_2 CI6_1 CI7_1	83.33 (17.56)	2.03 (0;07)
3.	ca. 200 word types	3	CI1_1 CI4_3 CI5_2	168.67 (16.07)	2;08 (0;07)
4.	ca. 350 word types	5	CI1_2 CI2_1 CI4_4 CI8_2 CI9_1	352.40 (69.76)	2;11 (0;07)
5.	> 500 word types	3	CI3_1 CI8_3 CI9_2	571.33 (43.32)	3;00 (0;09)

Table S2. Precise matching of children with ABI with children with NH.

	Cumulative vocabulary size ABI group	Number of CI files matched	Child IDs	Mean cumulative vocabulary size (SD)	Mean chronological age (SD) (years;months)
1.	< 50 word types	3	NH1_1 NH2_1 NH3_1	41.33 (8.32)	1;04 (0;01)
2.	ca. 100 word types	3	NH4_1 NH5_1 NH6_1	95.67 (25.38)	2;00 (0;00)
3.	ca. 200 word types	3	NH7_1 NH8_1 NH3_2	181.33 (22.31)	2;00 (0;00)
4.	ca. 350 word types	3	NH1_2 NH2_2 NH9_1	341.00 (14.18)	2;00 (0;00)
5.	> 500 word types	N.A.	N.A.	N.A.	N.A.
Chronological age and hearing age are similar in children with NH					