AJSLP

Research Article

Intraword Variability in Children With Cochlear Implants: The Long-Term Development up to 5 Years of Age and a Comparison With Children With Normal Hearing

Jolien Faes[a] and Steven Gillis[a]

Purpose: This study evaluates intraword or token-to-token variability in the spontaneous speech of Dutch-speaking children with cochlear implants (CIs) longitudinally up to 5 years of age in comparison with intraword variability in age-matched peers with normal hearing (NH).
Method: Spontaneous speech samples of 9 children with CI were collected longitudinally up to age 5. The data of the NH control group consisted of cross-sectional recordings. Children's word productions were categorized into 4 response types of the variability score (consistent correct, consistent incorrect, variable with hits, variable with no hits), and the proportion of whole-word variation (PWV) was calculated.
Results: PWV was high in both groups of children but decreased with age. All response types of the variability score appeared in both groups. Children with CI were significantly more variable than their peers with NH up to age 4, but this difference has disappeared by age 5. Longer words had a higher PWV and were more often consistent incorrect and variable.
Conclusions: Intraword variability was characteristic of children with CI's spontaneous speech productions as it was in children with NH, and a similar factor (word length) affected variability in production. Group comparisons showed higher rates of intraword variability in children with CI, but they seemed to catch up with their peers with NH by age 5.

A well-known characteristic of children's early word productions is their variability; children produce a particular word in different ways on different occasions. For instance, in a recording of the Dutch-speaking boy Maarten at age 1;11.08 (years;months.days), the child produced the proper name "Dominiek" /dominik/ in at least five different ways: [dɔmənik], [dɔmnik], [mik], [əmik], [mənik], [dənik] (Gillis, 2000). This type of intraword or token-to-token variability in typically developing children with normal hearing's (NH) early word productions has received ample attention in the literature (e.g., Sosa, 2015). Whether this phenomenon is equally frequent and develops in a similar fashion in children with congenital hearing impairments who received a cochlear implant (CI) in comparison with their peers with NH has hardly been investigated. Hence, little is known about intraword variability in the productions of children with CI, and the longitudinal development of this variability has not been studied yet. In the present article, the longitudinal development of intraword variability is studied in Dutch-speaking early implanted children with CI and compared to that of their peers with NH.

In what follows, the measures of intraword variability and the relevant literature on intraword variability in children with NH and children with CI will be reviewed.

Definition and Measures of Intraword Variability

Intraword variability denotes the phenomenon that multiple productions of a particular adult target word (i.e., the target) differ at a phonemic level from one another in a child's renditions (i.e., replicas) of that target word at a particular age, regardless of the accuracy of those different productions. For instance, if the Dutch adult target word /buk/ ("*boek,*" Eng. "book") is rendered once as /bu/

[a]Computational Linguistics & Psycholinguistics, University of Antwerp, Belgium

Correspondence to Jolien Faes: jolien.faes@uantwerpen.be

Editor: Krista Wilkinson
Associate Editor: Cynthia Cress

Received August 1, 2016
Revision received February 27, 2017
Accepted November 28, 2017
https://doi.org/10.1044/2017_AJSLP-16-0130

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

and once as /buk/, the child produces two different replicas of the target. Intraword variability was frequent in the speech of children with NH from word onset onward (Ferguson & Farwell, 1975; Holm, Crosbie, & Dodd, 2007; Macrae, 2013; McLeod & Hewitt, 2008; Sosa, 2015; Sosa & Stoel-Gammon, 2006, 2012; Taelman & Gillis, 2002). Even though intraword variability decreased with age (Holm et al., 2007; Sosa, 2015), it was still found in children with NH between 5 and 10 years of age (de Castro & Wertzner, 2011).

Two measures of intraword variability have been commonly used in the literature. First of all, intraword variability has been assessed by the proportion of whole-word variation (PWV) measure (Ingram, 2002). PWV computes the proportion of variability per word, with one indicating complete variability, meaning that each instance of a particular target word is pronounced differently (all replicas differ at a phonemic level), and zero indicating consistent production, which means that all instances of a particular target word are pronounced in the same way (all replicas are identical at a phonemic level). In other words, the degree of variability per target word is assessed. Multiple different productions (replicas) of a particular target word receive a higher PWV score than multiple similar or phonemically identical productions (replicas) of that word.

A second method that has been used to assess intraword variability categorizes the child's replicas per attempted target word (Holm et al., 2007; McLeod & Hewitt, 2008; Sosa, 2015): the variability score, also referred to as inconsistency score (Holm et al., 2007). The variability score categorizes each target word, or the adult form, relative to the child's renditions of that target. Four possibilities are usually distinguished: (a) all the child's productions are similar and correct (consistent correct); (b) all the child's productions are similar, but incorrect (consistent incorrect); (3) the child's productions differ, but at least one is correct (variable with hits); and (d) the child's productions differ, and none of them is correct (variable with no hits). In contrast to the PWV measure, the variability score takes the accuracy of the child's production into account, as it distinguishes between correct and incorrect production.

Intraword Variability in Children With NH

Especially at younger ages and in earlier stages of lexical development, children with NH exhibit elevated rates of intraword variability. For instance in binomial classifications, in which all replicas are evaluated for phonemic similarities and differences, variability mean scores were 78% for English-speaking children at a mean age of 2;04 (Macrae, 2013) and 74% for Dutch-speaking children between ages 1;00 and 2;09 (Taelman & Gillis, 2002). PWV scores were relatively high between 1 and 4 years of age (Sosa, 2015). For instance in spontaneous speech, PWV scores were approximately 46% at age 1;00 and 37% for the same group of children at age 2;00 (Sosa & Stoel-Gammon, 2006). Thus, the intraword variability is relatively high, but decreases with age. Similarly, Sosa and Stoel-Gammon (2012) reported that almost half of the productions of each target word are variable (49%) in children aged 2–2;06. For words with consonant clusters, the mean PWV score was about 54% between 2;06 and 3 years of age (McLeod & Hewitt, 2008). In contrast, Holm et al. (2007) found a much lower PWV percentage. For instance between 3;00 and 3;05, the mean PWV score was 13%, and between 6;00 and 6;11 the mean PWV was only 2.50%. Even though these figures also show a decrease with age, they are much lower than the percentages that are generally reported in the literature. In any case, it can be assumed that intraword variability, as measured by PWV, is high in children with NH from word onset up to approximately 4 years of age.

With respect to the variability score, both McLeod and Hewitt (2008) and Sosa (2015) showed that all four response types (consistent correct, consistent incorrect, variable with hits, and variable with no hits) are common in typical child language up to age 4. In contrast, Holm et al. (2007) indicated that the variable with no hits response type was rare in children with NH's productions from age 3 up to age 7. As for PWV, the results of Holm et al. (2007) differed from the other literature.

Several factors influence the likelihood of variable production of target words. A child-related factor is obviously age. Overall, intraword variability decreases with age. However, Ferguson and Farwell (1975) showed that children's first word productions were accurate, became less accurate with age, and eventually became accurate again. In other words, accuracy exhibited a U-shaped curve. It may well be that variability also follows this kind of developmental trend. High-accuracy rates go hand-in-hand with low variability, but the inverse is not necessarily the case. Low accuracy in production may be due to higher variability of children's productions, but it may also result from consistently using the same incorrect word form. In the latter case, low accuracy is accompanied by low variability.

Another child-related factor is vocabulary size. Intraword variability was lower in children with larger vocabularies (Macrae, 2013; Sosa & Stoel-Gammon, 2012). Also characteristics of the adult target word appeared to influence the likelihood of variable production. For instance, target words including late-acquired segments (e.g., fricatives or consonant clusters) or longer target words in terms of the number of segments or syllables have been shown to have higher variability rates than earlier segments or shorter target words (Leonard, Rowan, Morris, & Fey, 1982; Macrae, 2013; McLeod & Hewitt, 2008; Sosa, 2015; Sosa & Stoel-Gammon, 2012). Thus, the phonological complexity of the adult target word also influences the intraword variability in children's production. In this study, the impact of the target word's length, in terms of the number of syllables it contains, will be analyzed as well as the effect of age on variability.

Most studies of intraword variability in which PWV and the variability score are used were mainly restricted to English-speaking children with NH. But their methodologies varied from the point of view of the study design (longitudinal [McLeod & Hewitt, 2008; Sosa & Stoel-Gammon, 2006] vs. a single point in development [Holm et al., 2007; Sosa, 2015; Sosa & Stoel-Gammon, 2012]), the type of speech studied (spontaneous speech [McLeod & Hewitt, 2008; Sosa & Stoel-Gammon, 2006, 2012] vs. picture-naming tasks [Holm et al., 2007; Sosa, 2015]), and the target words (determined in advance [Holm et al., 2007; McLeod & Hewitt, 2008; Sosa, 2015; Sosa & Stoel-Gammon, 2012] or not [Sosa & Stoel-Gammon, 2006]). Also the computation of the relevant measures differed: In some studies, vowels were excluded from the variability counts (Sosa & Stoel-Gammon, 2006, 2012); in others, vowels were included (Holm et al., 2007; McLeod & Hewitt, 2008; Sosa, 2015). Despite these methodological differences, the general conclusions were highly similar: Intraword variability was high in the early stages of lexical language development of children with NH and decreased with age. Moreover, variability was higher for longer adult targets.

Intraword Variability in Children With CI

Thus far, few studies have investigated intraword variability in the speech of children with CI. Ertmer and Goffman (2011) studied intraword variability in six English-speaking children with CI in comparison with six peers with NH at a mean age of 4;00. In a toy-naming task, each child was asked to name each of the 60 toys three times. The children with CI were implanted before the age of 3;00 and had at least 2 years of device experience at the time of testing. The children with NH were matched on chronological age. All target words were mono- and disyllables—except for one word with three syllables "banana"—and the phonemic complexity target words' onsets varied. More specifically, the target words differed in the manner of articulation of the first consonant, and target words with, for instance, a word initial fricative were considered to be more complex than those with an initial stop. Results showed that the PWV score was higher for children with CI than for their peers with NH and that both groups of children had higher PWV scores in target words with word initial fricatives, affricates, and liquids (Ertmer & Goffman, 2011). Thus, the influence of different phonological features on intraword variability was similar in both groups of children.

Moreno-Torres (2014) investigated intraword variability in the spontaneous speech of eight Spanish-speaking children with CI and three peers with NH. The children with CI were followed longitudinally as part of a larger study, and the data of the children with NH were extracted from the CHILDES database (MacWhinney, 2000). Only one speech sample of each CI and NH child was selected for the analyses, which was the sample in which the child's mean length of utterance was closest to 1.2. The mean chronological age of the children with CI was 2;09 (*SD* = 0;04), and the mean age at implant activation was 1;04 (*SD* = 0;03), hence approximately a year and a half of device experience. A total of 159 target words were analyzed of children with CI and 72 of children with NH. The first two replicas of these target words were selected. Thus, the total number of analyzed tokens was relatively low (159 × 2, 72 × 2). The results showed that the PWV was considerably higher in children with CI (.63) than in children with NH (.15). Moreover, there were differences between both groups of children regarding the response types of the variability score. The children with CI had mainly variable with no hits and consistent incorrect responses, whereas children with NH had mainly consistent incorrect and consistent correct responses. In other words, Moreno-Torres (2014) found that children with NH's productions were more consistent and less variable than those of children with CI. Furthermore, the number of syllables in the target words had an effect on intraword variability; variability was slightly higher in longer words. This effect was established in the NH group, but not in the CI group.

Thus, Moreno-Torres (2014) and Ertmer and Goffman (2011) arrived at different conclusions regarding the effect of phonological complexity on intraword variability. The former did not find an effect of word length on variability in children with CI, whereas in the latter study such an effect was established. A possible explanation for these discrepant findings may be found in differences in the length of the words children with CI and NH use. Schauwers, Taelman, Gillis, and Govaerts (2008) showed that, between 2;00 and 2;06, children with CI produced shorter target words, which may explain the difference between children with NH and children with CI regarding the effect of a word length in Moreno-Torres (2014). He found that longer target words were more variable in children with NH. But, if a target word's syllable length tends to be shorter—as was the case in children with CI (Schauwers et al., 2008)—it may very well be that the effect was not visible yet in children with CI. In addition, Moreno-Torres (2014) considered only segmental differences for computing the variability measure and did not consider differences in the number of syllables of the children's productions. In other words, if one or more of a target word's syllables was truncated, this deletion was not reflected in the variability measure. This procedure may have influenced the results reported by Moreno-Torres (2014) in that the measure that he used did not account for the effect of word length on word productions.

Ertmer and Goffman (2011) and Moreno-Torres (2014) found higher rates of intraword variability in children with CI than in children with NH. In addition, Ertmer and Goffman (2011) showed that the phonological complexity of a target word affects intraword variability not only in children with NH but also in children with CI. In contrast, Moreno-Torres (2014) found an effect of word length on intraword variability in children with NH, but not in children with CI. In this study, children with NH and children with CI acquiring Dutch will be studied.

The Present Study

In this study, intraword variability is analyzed in the spontaneous speech of Dutch-speaking children with CI longitudinally. To the best of our knowledge, no information is available on intraword variability of Dutch-speaking children with CI thus far. We address the following specific research questions:

1. How does intraword variability in children with CI develop longitudinally in the preschool years?
2. Does age at implant activation influence intraword variability in children with CI?
3. Does the length of the target word affect intraword variability in children with CI?
4. Do children with CI have more intraword variability in the preschool years than children with NH? And, is the effect of word length similar in both groups of children?

Method

Participants

The data used in this study are part of the CCLC (CLiPS Child Language Corpus). CCLC contains audio and video recordings of the spontaneous speech of children with CI and children with NH. The corpus was collected and processed as part of several past research projects. The research protocols, details about the participants, data collection, transcription, and annotation were fully described in Hide (2013), Molemans (2011), Schauwers (2006), van den Berg (2012), and Van Severen (2012).

All participating children were monolingual Dutch and lived in Flanders, in the northern part of Belgium. Their parents were monolingual speakers of Dutch with no reported hearing deficits.

Nine children with CI were followed longitudinally from the moment their device was activated up to age 5;00. Monthly recordings of spontaneous dyadic interactions were available up to 30 months post implant activation, and thereafter, yearly data were available at ages 3;00, 4;00, and 5;00. The children had a congenitally profound hearing loss with a mean unaided pure-tone average (PTA) of 112.56 dB HL (*SD* = 9.12) in the better ear. No other apparent health or developmental problems were reported during data collection. All children received a Nucleus-24 implant before age 1;08. The mean age at implantation was 1;00 (*SD* = 0;05), and the mean age at device activation was 1;01 (*SD* = 0;05). After implantation, the mean PTA improved to 32.33 dB HL (*SD* = 7.11) at age 5;00. All children used oral communication with a limited support of lexical signs. Three children received a second CI within the period studied. Detailed information and individual data can be found in Table 1.

A cross-sectional group with a total of 42 children with NH participated as a control group: ten 2-year-olds (mean = 2;00, *SD* = 0;01), nine 3-year-olds (mean = 3;00, *SD* = 0;01), twelve 4-year-olds (mean = 4;00 months, *SD* = 0;01), and eleven 5-year-olds (mean = 5;00, *SD* = 0;01). These children were recorded only once in spontaneous dyadic interactions at their homes.

Table 1. Characteristics of children with cochlear implants.

			Age activation		
ID	**PTA unaided**	**PTA CI**	**Age 1st CI**	**1^{e} CI**	**Age 2nd CI**
S1	120	35	1;01	1;03	—
S2	120	27	0;06	0;08	4;08
S3	115	25	0;10	1;00	—
S4	113	42	1;06	1;07	—
S5	93	32	1;05	1;06	—
S6	120	37	0;09	0;10	—
S7	117	23	0;05	0;06	1;03
S8	112	42	1;07	1;09	—
S9	103	28	0;09	0;10	1;11
Mean	112.56	32.33	1;00	1;01	2;07
SD	9.12	7.11	0;05	0;05	1;09

Note. Ages are represented in years;months. PTA = pure-tone average (in dB HL); CI = cochlear implant.

Data Collection and Data Transcription

Audio and video recordings of spontaneous speech samples of children interacting with their caregiver(s) were collected for 60–90 min at the children's homes. The observation sessions were not structured in any way; the parents were asked to interact with their children as they usually do. In order to keep the transcription time within reasonable limits, a selection was made from the original recordings. After each recording, the researcher who had made the recording selected 20 min of recording, aiming at including delineated sequences of interaction in which the child was the most vocally active. An episode on a certain topic was never interrupted in the selection process, so that only completed interactions or episodes were included. Long pauses and noisy passages were avoided.

Each 20-min selection was transcribed using CHILDES' CLAN program according to the CHAT conventions (MacWhinney, 2000). Children's productions were transcribed orthographically, and a broad phonemic transcription was made in DISC symbols. In (1) an example is provided for the Dutch word "boek" (Eng. "book"), produced by a child as [bu]. The first line, the speaker tier (*CHI), presents the orthographic transcription ("boek"). The second line, the % pho tier, contains the phonemic transcription of the actual child production in DISC symbols: /bu/ is the phonemic transcription of the child's actual production. The phonemic transcription of the target word or the adult equivalent of the child's production is added on the %ohp tier (/buk/ in example "1"). The %ohp tier represents the phonemic transcription (in DISC symbols) of the standard pronunciation of the target word, which was retrieved from the lexical database Fonilex. Fonilex is "a pronunciation database containing the [broad phonemic] transcription of the most frequent word forms of Dutch as spoken in Flanders" (Mertens, 2001).

(1) *CHI: boek
%pho: bu
%ohp: buk

For the identification of words, the procedure proposed by Vihman and McCune (1994) was used. In order to be counted as a word, a child's production had to meet a number of criteria relative to its shape, its context of use, and its relation to other vocalizations. The criteria based on vocalization shape involved, for instance, a complex match of the vocalization with the target production, such that more than two segments of the child production matched with those in the target form. The criteria based on context covered, for instance, the identification of the vocalization as a word by the mother. The criteria based on the relation to other vocalizations involved, for instance, the absence of inappropriate use, such as the vocalization was only used in plausible contexts.

For each child at each age, the replicas of each target word token were listed. A total of 24,942 target word tokens were available (NH: 9,301; CI: 15,641). Target word tokens with only one child replica at a particular age were excluded as no intraword variability score can be derived from a single word token (Ingram, 2002). This resulted in a total of 16,640 target word tokens (NH: 6,942; CI: 9,698). Next, the suggestion of Ingram (2002) was followed: Because the PWV measure is sensitive to repetitions, a limit has to be defined on how many repetitions are allowed for the computation of PWV. If there is only one production, variation is impossible, and these cases have been excluded. Ingram put the upper bound to three instances or tokens of a particular adult target. We have extended this limit to 10 tokens of a target. If a target word type had more than 10 child realizations (replicas), the target word was left out of the data analyses, which eventually resulted in a total of 14,383 target word tokens (CI: 8,378 and NH: 6,005).

The reliability of two aspects of the transcriptions was checked: the identification of the target words and the phonemic transcription of children's productions. For the reliability of the identification of the target words, approximately 25% of the data were reannotated by a second transcriber (interrater reliability). The percentage of agreement on the identified target words equaled 81.38%.

Approximately 10% of the data were retranscribed in order to check inter- and intrarater reliability of the phonemic transcriptions.[1] The percentage of interrater agreement for NH speech samples was 63.69% for a phoneme-to-phoneme comparison. If only articulatory features were considered, percentages of agreement equaled 81.14% for consonant place of articulation, 78.70% for consonant voicing, and 81.03% for consonant manner of articulation and equaled 75.69% for vowel place of articulation, 77.23% for vowel height, and 81.94% for vowel roundedness. Intrarater reliability for NH speech samples was 81.51% for the phoneme-to-phoneme comparison. Considering articulatory features, intrarater reliability equaled 92.08% for consonant place of articulation, 87.50% for consonant voicing, 91.72% for consonant manner of articulation, 88.25% for vowel place of articulation, 88.93% for vowel height, and 90.85% for vowel roundedness.

For the CI corpus, the percentage of agreement was checked for interrater reliability and equaled 81.63% for the phoneme-to-phoneme comparison. Percentages of agreement of articulatory features were 82.90% for consonantal place of articulation, 85.70% for consonant manner of articulation, 73.50% for vowel place of articulation, and 84.54% for vowel height.

In addition, Kappa scores were calculated in order to account for possible influence of chance (Cucchiarini, 1996). For the NH speech samples, Kappa scores were .60 for interrater reliability and .80 for intrarater reliability of the phoneme-to-phoneme comparison. These scores can be interpreted as on the edge of "moderate" to "substantial" and as on the edge of "substantial" to "almost perfect" agreement, respectively (Landis & Koch, 1977). For the CI speech samples, the Kappa score is .87 for interrater reliability of the phoneme-to-phoneme comparison, which can be interpreted as "almost perfect" (Landis & Koch, 1977). The discrepant reliability scores for the NH and the CI corpus can—at least partly—be explained by the specific characteristics of the transcribers and more precisely their regional background. For the CI corpus, the two transcribers came from the same region, the Brabantic area, whereas for the NH corpus, the three transcribers came from three different regions: one from the province of Limburg, one from the province East Flanders, and the third one from the province North Holland (Randstad) in the Netherlands. Research has shown that transcribers' regional background has a pervasive though predictable influence (Kloots, Coussé, & Gillis, 2006).

Variability Measures

For each child at each age, the number of child replicas, the number of different child replicas, and the number of correct replicas for each adult word attempted by the child were calculated. Two measures of intraword variability were computed: the PWV and the variability score. For these measures, all segmental information was included, such that both consonantal and vocalic differences were considered. The number of syllables was counted and served as an index of word length.

The first measure, PWV, was calculated as the ratio of the number of distinct child replicas of a particular adult target over the number of child tokens (Ingram, 2002). For instance, if a child produced the Dutch adult target /brur/ "broer" (Eng. "brother"), once as /bur/ and once as /bu/, the child produced two different replicas of the target /brur/. If the child produced each replica only once, then there were two tokens of the attempted adult word, and PWV equaled 2/2 = 1.0. Alternately, the child attempted to produce the adult target /pus/ "poes" (Eng. "cat") six times, three times as /pupu/, twice as /pu/, and once as /pus/. In this example the adult target /pus/ resulted in three distinct replicas (/pupu/, /pu/, /pus/), and the total number of

[1]The reader is referred to Molemans (2011, pp. 41–77) for a description of the retranscription procedure and to Molemans (2011, pp. 90–93) for the confusion matrices, detailing the overlaps and discrepancies between the original transcription and retranscription.

child tokens equaled six. Consequently, PWV equaled 3/6 = .5. Ingram (2002, p. 720) argued that if the child produces an adult target word such as /buk/ “boek” (Eng. “book”) consistently as /bu/, there is only one child replica, and hence zero different replicas. Consequently PWV equaled 0/5 = 0. This indicates that the child produces that word consistently. It can readily be seen that the resulting PWV scores range between zero and one.

For each target word, the variability score was determined as well. The variability score comprised four categorical response types:

- Consistent correct: The child’s replicas are phonemically identical and match the target.
- Consistent incorrect: The child’s replicas are phonemically identical but do not match the target.
- Variable with hits: The child’s replicas are phonemically variable, but at least one matches the target.
- Variable with no hits: The child’s replicas are phonemically variable, with none matching the target.

Statistical Analysis

Statistical analyses were performed in R by means of multilevel models (R Core Team, 2013). Multilevel models were most appropriate for our data analyses, as our data were structured hierarchically in three levels: At the lowest level, there were the particular target words produced; at the next level, productions of target words were observed at a particular age or observation session, which were in their turn nested within individual children (third level). In other words, there was some variation in our data set resulting from the nesting of variables at different levels. Multilevel models consist of two parts: a random part and a fixed part. The random part considers the variation in the data caused by, among other things, the nesting of variables (Baayen, 2008; Woltman, Feldstain, MacKay, & Rocchi, 2012). In the fixed part, the predicting variables are modeled. A model was constructed in an iterative way: Random and fixed effects were added to the model one after the other, and with each addition, it was tested with an analysis of variance if the addition of the variable yielded a better model fit. If the model fit improved significantly, the variable was included in the final model that will be reported in the Results section.

Two different measures of variability were examined. First of all, the PWV score was analyzed in a multilevel model. Second, the variability score was analyzed using a logistic regression in a multilevel model. In other words, the likelihood of each response type (as compared to all other response types) was examined. These results are expressed in logits.

For each variability measure (PWV, variability score), there were two parts in the analyses. In the first part, the development of intraword variability in children with CI was traced longitudinally. For this purpose, the monthly data from word onset (median = 1;06, range 1;03–1;11) up to age 2;06 and the yearly data at ages 3;00, 4;00, and 5;00 were considered. Because these data were observations of the same children over time, they can be included in a single model. In the second part, the children with CI were compared with children with NH. The data of the children with CI were the same as the data in the first analysis. But data collection of the NH was cross-sectional, which does not warrant including all the data in a single model. Therefore, the children with CI and the age-matched children with NH were compared separately at ages 2;00 (range of the NH: 1;11–2;01), 3;00 (range of the NH: 2;10–3;04), 4;00 (range of the NH: 3;09–4;03), and 5;00 (range: 4;11–5;03).

For the longitudinal analyses, the intercept was set at age 1;10 as we have the largest amount of data at this age, and the age of implant activation was centered at the mean (1;01). In each analysis, the fixed effects were age (age), the number of syllables of the target word (syllable length), and the age at implant activation (CI activation). Interactions between those variables as well as quadratic and cubic age effects were tested and included into the final model if they significantly improved the model fit. Because language development has been shown to be nonlinear (e.g., Ferguson & Farwell, 1975), higher-order age effects were added to the analyses in order to investigate if the development with age was linear, quadratic, or cubic within the studied period.

For the cross-sectional analyses, the fixed effects were hearing status (hearing status) and the number of syllables (syllable length), and interactions between hearing status and syllable length were included if that yielded a better model fit. In order to model the variation in the data, we allowed random intercepts and slopes in the longitudinal analyses and random intercepts for the cross-sectional analyses. A significance level of $p < .05$ was set.

Results

Intraword Variability in Children With CI

The first part of the analyses presents the results of the longitudinal analyses of children with CI for PWV and the variability score, respectively. The fixed effect results for PWV can be found in Table 2, and those for the variability score are in Table 3. The results in Table 3 are expressed in logits. For the sake of familiarity, logits are converted to probabilities when they are given in the presentation of the results.

Longitudinal Development With Age

Children with CI’s productions exhibited intraword variability. At the intercept (i.e., 22 months of age), the estimated average PWV score equaled 0.30, which is well above zero ($p < .001$ for the intercept; see Table 2). Intraword variability decreased significantly with age ($p < .001$). The estimated PWV values were plotted in Figure 1.

For the variability score, the likelihood of consistent correct, consistent incorrect, variable with hits, and variable with no hits responses was 7.59%, 46.26%, 11.41%,

Table 2. Fixed effect results of the longitudinal analysis of children with cochlear implants–proportion of whole-word variation.

Fixed effects	Estimate (*SE*)
Intercept	0.30 (0.02)***
Age	–0.01 (0.00[a])***
Syllable length	0.02 (0.01)**
Syllable length × Age	< –0.01 (0.00[a])**
CI activation	< 0.01 (0.00[a])

Note. *SE* = standard error; CI = cochlear implant.

[a]A standard error of 0.00 indicates a value lower than 0.01.

**$p \leq .01$. ***$p \leq .001$.

and 32.74%, respectively, at the intercept (see Table 3). These percentages do not add up to 100%, as the likelihood of each response type was estimated relative to the other three response types, but they give a good indication of the distribution of the different response types. In Figure 2, the development with age of each response type was plotted. The likelihood of consistent correct responses significantly increased with age, whereas the likelihood of consistent incorrect and variable with no hits responses significantly decreased with age ($p < .001$ in all analyses). The likelihood of variable with hits responses first significantly increased with age ($p < .001$) but decreased from approximately the age of 45 months (quadratic effect of age, $p < .001$).

Effect of Age at Implantation

The age at CI activation was only significant for the variability score. With respect to PWV, the effect of CI activation was not significant ($p > .05$; Table 2), meaning that the PWV scores were similar for children with CI with earlier and later implant activation. However, as measured by the variability score, later CI activation was related to a lower correctness. Children with CI with later implant activation were significantly less likely to have consistent correct ($p < .01$) and variable with hits ($p < .05$) responses and were significantly more likely to have consistent incorrect ($p < .01$) and variable with no hits ($p < .001$) responses. However, the difference between children with earlier and later implant activation became less pronounced with age for the categories consistent correct and variable with no hits responses as can be inferred from the significant interaction between the variables age and CI activation in Table 3.

Effect of the Length of the Target Word (Syllable Length)

In the word productions of children with CI, intraword variability varied relative to word length; longer words were more variable. The PWV was higher for longer target words ($p < .01$; Table 2), and the likelihood of the response types in which correct responses appear was lower with increasing syllable length of the target word. Longer target words were less likely to be consistent correct and variable with hits and were more likely to be consistent incorrect and variable with no hits ($p < .001$; Table 3).

Nevertheless, these patterns changed with age, as shown by some significant interactions between syllable length and age. The increase of PWV with increasing syllable length became less pronounced over time ($p < .01$). The decrease of consistent correct responses in longer target words became smaller with age as well ($p < .05$), but the effect of syllable length on consistent incorrect and variable with hits responses increased with age ($p < .01$ and $p < .05$).

Comparison of Children With NH and Children With CI

The second part of the analysis concerns the cross-sectional comparisons of children with CI and age-matched

Table 3. Fixed effect results of the longitudinal analysis of children with cochlear implants–variability score.

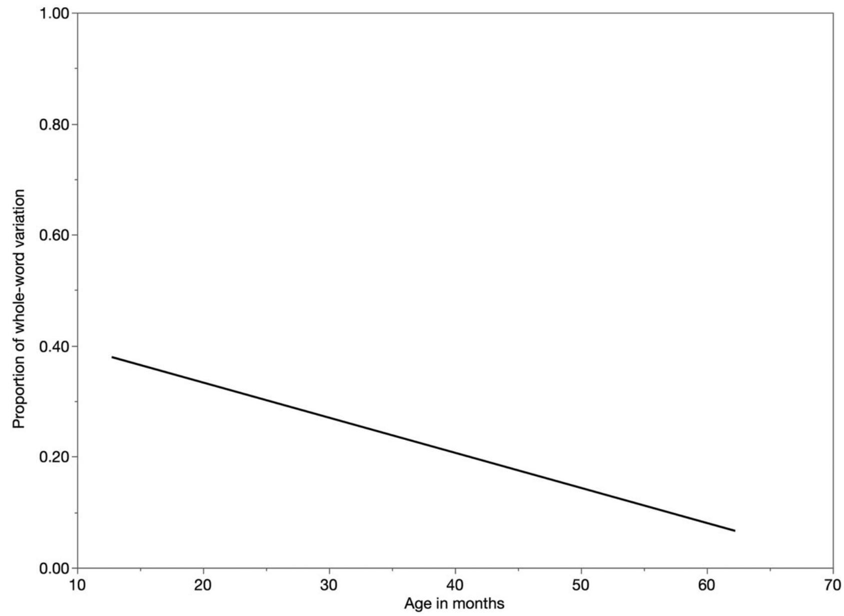
Fixed effects	Estimates and *SE*s	Consistent correct	Consistent incorrect	Variable with hits	Variable with no hits
Intercept	Estimate	–2.50***	–0.15***	–2.05***	–0.72***
	SE	0.23	0.09	0.23	0.16
Age	Estimate	0.09***	–0.05***	0.14***	–0.08***
	SE	0.01	0.01	0.02	0.01
Age × Age	Estimate			–0.01***	
	SE			0.00[a]	
Syllable length	Estimate	–0.37***	0.36***	–0.74***	0.33***
	SE	0.05	0.04	0.07	0.05
Syllable length × Age	Estimate	0.01***	0.01**	–0.01*	0.00[a]
	SE	0.00[a]	0.00[a]	0.01	0.00[a]
CI activation	Estimate	–0.14**	0.04**	–0.09*	0.10***
	SE	0.04	0.01	0.04	0.03
CI activation × Age	Estimate	0.06***	0.00[a]	0.00[a]	0.00[a,*]
	SE	0.00[a]	0.00[a]	0.00[a]	0.00[a]

Note. *SE* = standard error; CI = cochlear implant.

[a]A standard error of 0.00 indicates a value lower than 0.01.

*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

Figure 1. Development of proportion of whole-word variation with age for children with cochlear implants—predicted values.



children with NH. First, the results for PWV are reported, and then those for the variability score. Table 4 displays the fixed effect results of PWV. These results are graphically represented in Figure 3. With respect to the variability score, the tables (in logits) are provided in Appendices A–D.

Intraword Variability in Children With CI and NH: Age Comparisons

Intraword variability decreased with age in both groups of children. PWV decreased from .60 to .22 in children with NH and from .45 to .05 in children with CI (see Table 4; Figure 3). The response types of the variability

Figure 2. Development of the different response types of the variability score with age for children with cochlear implants—predicted values in logits.

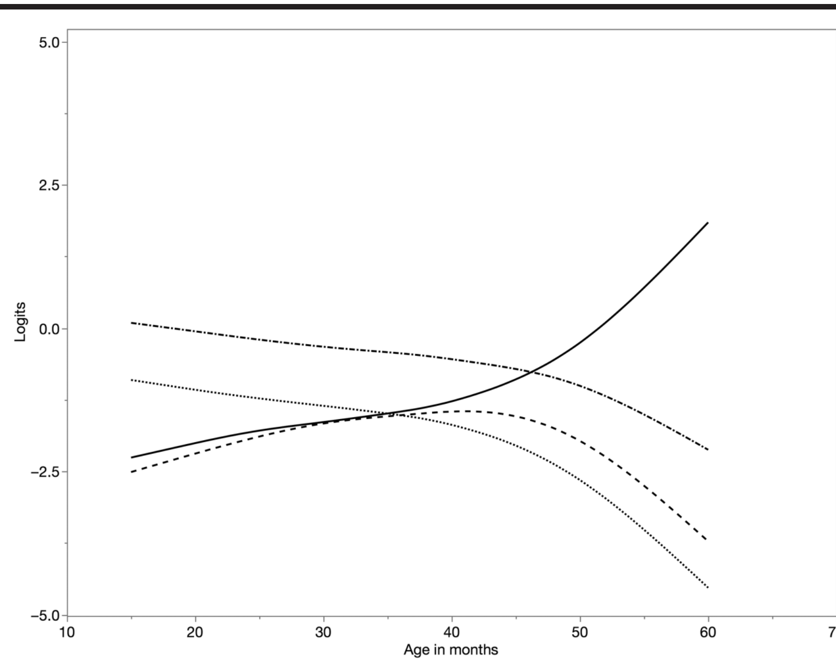


Table 4. Fixed effect results of proportion of whole-word variation scores–cross-sectional comparisons between normal hearing and cochlear implant.

	2;00	3;00	4;00	5;00
Fixed effects	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)
Intercept	0.60 (0.03)***	0.38 (0.04)***	0.24 (0.03)***	0.22 (0.04)***
Hearing status CI	–0.15 (0.05)**	0.06 (0.05)	0.12 (0.05)**	–0.17 (0.06)**
Syllable length	0.12 (0.02)***	0.06 (0.03)*	0.02 (0.02)	–0.02 (0.02)
Syllable length × Hearing status CI	0.02 (0.05)	0.04 (0.04)	0.08 (0.03)**	0.02 (0.02)

Note. Ages are represented in years;months. *SE* = standard error; CI = cochlear implant.
*$p \le .05$. **$p \le .01$. ***$p \le .001$.

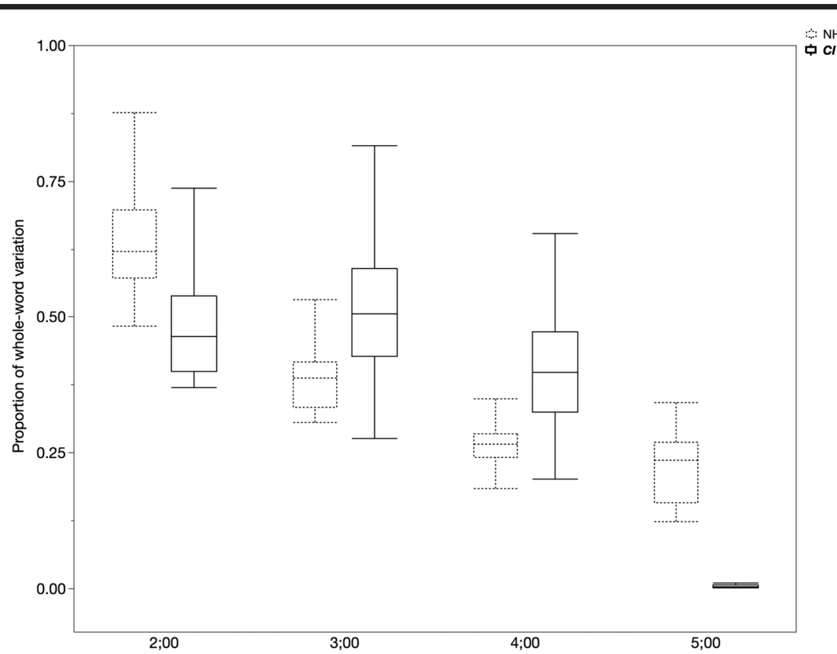
score developed toward more consistency and more correctness (Appendices A–D). The amount of intraword variability differed in children with NH and CI, depending on their age.

At age 2;00, children with CI were less variable as compared to their peers with NH: The PWV and the likelihood of variable response types were significantly lower ($p < .001$ and $p < .01$, respectively; variable with hits: NH: 37.99% vs. CI: 21.42%). However, children with CI were also less correct as compared to peers with NH (consistent incorrect responses: CI: 25.35%, NH: 2.90%, $p < .01$). Because further analyses revealed that the target words of children with CI were significantly shorter (syllable length) than those of children with NH at age 2;00 (estimates: .46 in CI, .61 in NH [*SE* = .06], $p < .01$), this may be the explanation of lower variability in children with CI. By age 3;00, syllable length was similar in the target words of both groups of children (estimate: .56 [*SE* = .03], $p > .05$).

At age 3;00, PWV values were similar in both groups of children ($p > .05$). Nevertheless, correctness of production seemed lower in children with CI than in children with NH. Children with CI were less likely to have consistent correct responses (18.24%, $p > .01$) and variable with hits responses (34.52%, $p < .05$) as compared to children with NH (32.74% and 45.51%), whereas they were more likely to have variable with no hits responses (26,31%, $p < .001$) than children with NH (10.34%).

At age 4;00, PWV values were significantly higher in children with CI ($p < .01$), indicating that they were more variable in production. A similar picture appeared in the variability score. Even though children with CI and NH were equally likely to have variable with hits responses (35.43%,

Figure 3. Development of proportion of whole-word variation with age for children with cochlear implants (CI) and children with normal hearing (NH)—predicted values.



$p > .05$), children with CI were significantly less likely to have consistent correct responses (30.15% vs. 51.50%, $p < .01$) and significantly more likely to have consistent incorrect (14.06% vs. 9.11%, $p < .01$) and variable with no hits responses (17.80% vs. 4.03%, $p < .001$).

At age 5;00, children with CI were less variable as compared to their peers with NH. PWV values for children with CI were lower than those of their peers with NH ($p < .05$). In addition, children with CI's productions were less likely to be variable with hits (CI: 0.53%, NH: 33.18%) and variable with no hits (CI: 0.29%, NH: 2.77%) and more likely to be consistent correct (CI: 89.93%, NH: 52.00%).

Effect of Target Word Length: Group Comparisons

Longer target words were more variable in both groups of children, but the precise effect differed. With respect to PWV, variability was similar in both groups of children at ages 2;00 and 3;00; longer words were more variable ($p < .001$ and $p < .05$). At age 4;00, the syllable length had no effect on PWV in children with NH ($p > .05$), but a significant interaction between syllable length and hearing status ($p < .01$) showed that the effect was still present in children with CI. At age 5;00, the effect of syllable length had disappeared in both groups of children ($p > .05$).

With respect to the variability score, incorrect and variable response types were found to be more likely in longer target words (Appendices A–D). At age 2;00, this effect was less pronounced in children with CI. This can be related to their shorter target words at this age (Schauwers et al., 2008). However, by age 3;00, the effect of target word syllable length was similar in both groups of children. One exception was found at age 4;00; at this age, consistent correct responses were less likely in longer words in children with CI, but not in children with NH ($p < .05$).

Discussion

Longitudinal Development of Intraword Variability in Children With CI

Intraword variability decreased in children with CI with age. There did not appear to be longitudinal studies of children with CI's intraword variability so far, but the decrease of intraword variability with age is similar to the findings of Holm et al. (2007) and Sosa (2015) for children with NH. The relatively high PWV percentages found in this study are in the range previously reported in the literature on intraword variability in children with CI (Moreno-Torres, 2014) and children with NH (Macrae, 2013; Sosa, 2015; Sosa & Stoel-Gammon, 2006, 2012). With respect to the variability score, results showed that all four response types appear in Dutch-speaking children with CI in the preschool years. In the CI literature, only Moreno-Torres (2014) indicated that Spanish-speaking 3-year-old children with CI had mainly incorrect responses. In contrast, our results pointed out that Dutch-speaking children with CI's productions can be categorized in all four response types up to the age of 5, which is also similar to the profile of children with NH (McLeod & Hewitt, 2008; Sosa, 2015).

Differences Between Children With CI and Children With NH

Children with CI produced words that were less variable than their peers with NH at the age of 2. At the age of 3, PWV of both groups of children was similar. From then on PWV fluctuated, reflecting most probably the idiosyncrasies of the participants in a cross-sectional sample; at the age of 4, the PWV of children with CI was higher than that of children with NH, whereas at the age of 5, the opposite was true.

An explanation for the lower rate of variability at the age of 2 may be found in the difference in target word length of both groups of children. Children with CI produced shorter target words, that is, words with fewer syllables, as compared to their peers with NH at this age. Thus, their target words were less complex in the sense that they contain fewer syllables (see also Schauwers et al., 2008). As shorter targets tended to be produced less variably and children with CI produced shorter words, this may explain the fact that they end up being less variable than children with NH.

An explanation for the lower variability in children with CI than in children with NH at the age of 5 is less straightforward. This may be a consequence of an inherent limitation of the cross-sectional nature of the data of the children with NH. The CI speech samples were collected longitudinally, whereas the NH speech samples were cross-sectional. The data for each age group of children with NH were collected from different children, and these samples did not show continuity across ages. Hence, there was little information on the previous language skills of those 5-year-olds with NH. Is the overall lower consistency characteristic for all 5-year-olds, or is it an accidental characteristic of the 5-year-olds in the present sample? This may be the case, though enlarging the database is the only way to figure this out.

There is an additional factor that needs to be taken into account. The decrease of the PWV score suggested that the number of different replicas of a given target decreased over time. But a PWV score of 0 does not mean that children are producing a target word correctly. It simply shows that they are consistent in their production. In other words, 5-year-old children with CI produced target words in general less variably than peers with NH, but it may very well be that target words at this age are produced consistently incorrect. The PWV measure provided an indication of the child's consistency in rendering target words, but it did not provide any information on the accuracy of production.

For this purpose, a second measure of intraword variability was used: the variability score. Up to the age of 4, children with CI were less likely to hit the target and less likely to be consistent in comparison to peers with NH. The 2-, 3-, and 4-year-old children with CI scored lower for the category consistent correct and consistent incorrect. But the 5-year-old children with CI had a higher score for consistent correct and a comparable score for consistent incorrect. This seemed to indicate that children with CI do not only catch up with their peers with NH by the age of 5; they even seemed to outperform their peers with NH for correct productions. These results are highly promising for the perspectives of children with congenital sensorineural hearing loss. At a phonemic level, their word productions matched the accuracy of their peers with NH at the age of 5. This does not mean that there is a complete match at a phonetic level, because Verhoeven, Hide, De Maeyer, Gillis, and Gillis (2016) showed that the same children with CI from this study had reduced vowel spaces in comparison to peers with NH.

The Effect of the Length of Target Words

Variability increased when target words contained more syllables in both groups of children. To the best of our knowledge, the only study that reported on the effect of word length on intraword variability in children with CI was Moreno-Torres (2014). In that study, no effect of syllable length in children with CI was found. It should be noted that syllable omissions from target words were not considered as variable productions in Moreno-Torres (2014). This may at least partly explain the discrepant findings of his study and ours. Our results corroborated the finding that phonological complexity has an effect on intraword variability. For instance, Ertmer and Goffman (2011) found that target words with late acquired word initial consonants (e.g., fricatives) are more variable in children with CI. This study showed that another factor of phonological complexity has a similar impact on children with CI's variability: variability increased with increasing word length. Because Ertmer and Goffman (2011) studied only one point in development (4 years of age), they could not track a possible decline of the effect of phonological complexity. This study took a longitudinal approach and revealed that the impact of word length decreases in children with CI with age.

As for the group differences, the results suggested that the effect of word length was less pronounced at the age of 2 in children with CI than in their peers with NH. This may be due to the fact that children with CI produced shorter target words at this age. Furthermore, the effect of word length on intraword variability was limited in time; it has disappeared by the age of 3 in children with NH, but only by the age of 4 in children with CI.

The Effect of Age at Implant Activation in Children With CI

The age at CI activation did not affect the PWV scores. This may be surprising because effects of age at device activation have been reported in several speech and language domains (Connor, Craig, Raudenbush, Heavner, & Zwolan, 2006; Faes, Gillis, & Gillis, 2016; Leigh, Detmman, Dowell, & Briggs, 2013; Nikolopoulos, O'Donoghue, & Archbold, 1999). However, it may be explained by the small age range of CI activation. Children with CI in this study were all implanted before 20 months of age, and the difference between the child with the earliest CI activation and the child with the latest activation was only 15 months. However, when considering the intraword variability in relation to correctness by means of the variability score, results pointed out that children with later implant activation are more likely to have variable and consistent responses.

Limitations

There are some methodological considerations to take into account that limit the generalizability of the results. The group sizes were relatively small in both groups of children. So the results of this study should be considered with some care. A larger sample of both children with CI and children with NH could reveal more insight into the precise group and age effects. In addition, the data design was slightly different depending on the group. Longitudinal data were available for the children with CI, but not for children with NH. The cross-sectional data collection for children with NH is an inherent limitation of this study and may limit the interpretation of the group comparisons. As discussed earlier, the remarkable result that 5-year-olds with NH have higher variability rates than 5-year-olds with CI may be due to this data design. The group comparisons would benefit from a longitudinal sample of children with NH in order to seek out if this effect at age 5 is inherent to our data. Finally, the effect of age at implant activation in children with CI should also be considered with care, because the range of implant activation was relatively small (15 months, see higher).

Clinical Implications

The results suggested that the level of detail of phonological representations differed in children with CI and children with NH. It is very likely that children with CI's phonological representations of target words are less fine-grained, which may increase their variability and inaccuracy in production. A focus on the underlying factors that hamper the fine-tuning could enhance their production. First, the speech signal is degraded and noisier in children with CI as compared to that available in NH (Drennan & Rubinstein, 2008). This means that the input signal is less clear in children with CI, which may hinder fine-tuning of phonological representations. Moreover, the degraded signal may also restrict children's perception of their own productions, like auditory feedback of their own productions. As auditory feedback is also essential in fine-tuning articulation (Stoel-Gammon, 2011; Stoel-Gammon & Sosa, 2007), it may add to production inconsistency in children with CI, as hypothesized by Moreno-Torres (2014).

Second, children with CI were found to be less attentive to the ambient language (Houston & Bergeson, 2014; Houston, Pisoni, Kirk, Ying, & Miyamoto, 2003). Even when the speech signal would be similar to that in NH, children with CI are likely to miss some parts of the input, as they pay less attention to it. As such, they have fewer opportunities to fine-tune their phonological representations. Training on attention to speech and the speech signal might enhance children with CI's fine-tuning of phonological representations.

Third, even if perception and attention would be similar in both groups of children, children with CI were found to have difficulties in processing and storing the input and processing their own output (Kronenberger, Pisoni, Henning, & Colson, 2013; Nittrouer, Caldwell-Tarr, & Lowensthein, 2013; Pisoni, Kronenberger, Roman, & Geers, 2010). Thus, they seemed less efficient in decoding the speech signal into a phonological representation and in the storage of this information, and they seemed less efficient the other way as well: from a phonological representation to an actual production. These difficulties may affect the fine-tuning of their phonological representations, but they may also affect articulation, as children with CI's less efficient processing of their own output may create more variability in production. In other words, a focus on processing skills in clinical intervention may be beneficial for children with CI.

Speech and language therapists might also focus on longer words in intervention, given the fact that the effect of syllable length on variability persisted longer in children with CI as compared to NH children. Processing of more complex words, such as longer words, was found to be more difficult in all children irrespective of their hearing conditions (Faes et al., 2016; Gathercole, Willis, Emslie, & Baddeley, 1991). As a result, more complex words have less fine-grained phonological representations initially, which may explain larger variability in production. In children with CI, the effect of word length was more pronounced than in peers with NH. This may be explained by their less developed processing skills, which result in less fine-grained phonological representations. In a similar vein, Von Mentzer et al. (2015) showed that children with CI's phonological representation of words with consonant clusters are also less fine-grained. Words with consonant clusters are also seen as complex words, similar to longer words. Both are produced less accurately and more variably: Faes and Gillis (2017) showed that children with CI were initially less accurate in their production of consonant clusters, and children with CI produced longer words less accurately (Faes et al., 2016) and more variably (this study).

Future Research

Thus far, only the effect of age, the effect of age at implantation, and the effect of word length on the intraword variability of children with CI have been studied. For children with NH, also other factors were reported to affect intraword variability, including word frequency, the phonological neighborhood density of target words, and children's vocabulary size (Macrae, 2013; Sosa, 2015; Sosa & Stoel-Gammon, 2012). In future studies, it is essential to investigate the effect of those factors on intraword variability in children with CI. As such, insight in the effect of those factors on the variability of children with CI can adjust the speech and language therapy in children with CI and can reveal a better understanding of the role of input and auditory feedback in language development.

Acknowledgments

This work was supported by Grant G.0138.13 of the Research Foundation–Flanders (FWO) to S. Gillis. Our special thanks go to I. Molemans, K. Schauwers, R. van den Berg, and L. Van Severen for collecting and transcribing the data analyzed in this study.

References

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

Connor, C., Craig, H., Raudenbush, S., Heavner, K., & Zwolan, T. (2006). The age at which young deaf children receive cochlear implants and their vocabulary and speech-production growth: Is there an added value for early implantation? *Ear and Hearing, 27,* 628–644. https://doi.org/10.1097/01.aud.0000240640.59205.42

Cucchiarini, C. (1996). Assessing transcription agreement: Methodological aspects. *Clinical Linguistics & Phonetics, 10,* 131–155. https://doi.org/10.3109/02699209608985167

de Castro, M., & Wertzner, H. F. (2011). Speech inconsistency index in Brazilian Portuguese–speaking children. *Folia Phoniatrica et Logopaedica, 63,* 237–241. https://doi.org/10.1159/000323183

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,* 779–790. https://doi.org/10.1682/JRRD.2007.08.0118

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)

Faes, J., Gillis, J., & Gillis, S. (2016). Phonemic accuracy development in children with cochlear implants up to five years of age by using levenshtein distance. *Journal of Communication Disorders, 59,* 40–58. https://doi.org/10.1016/j.jcomdis.2015.09.004

Faes, J., & Gillis, S. (2017). Consonant cluster production in children with cochlear implants: A comparison with normally hearing peers. *First Language, 37,* 319–349. https://doi.org/10.1177/0142723717692631

Ferguson, C., & Farwell, C. (1975). Words and sounds in early language development. *Language, 51,* 419–439.

Gathercole, S., Willis, C., Emslie, H., & Baddeley, A. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics, 12,* 349–367. https://doi.org/10.1017/S0142716400009267

Gillis, S. (2000). Phonological development. In S. Gillis & A. Schaerlaekens (Eds.), *Child Language Acquisition. A manual for Dutch*. Groningen, The Netherlands: Martinus Nijhoff.

Hide, O. (2013). *Acoustic features of speech by young cochlear implant users*. Belgium: University of Antwerp.

Holm, A., Crosbie, S., & Dodd, B. (2007). Differentiating normal variability from inconsistency in children's speech: Normative

data. *International Journal of Language & Communication Disorders, 42,* 467–486. https://doi.org/10.1080/13682820600988967
Houston, D., & Bergeson, T. (2014). Hearing versus listening: Attention to speech and its role in language acquisition in deaf infants with cochlear implants. *Lingua, 139,* 10–25. https://doi.org/10.1016/j.lingua.2013.08.001
Houston, D., Pisoni, D., Kirk, K., Ying, E., & Miyamoto, R. (2003). Speech perception skills of deaf infants following cochlear implantation: A first report. *International Journal of Pediatric Otorhinolaryngology, 67,* 479–495. https://doi.org/10.1016/S0165-5876(03)00005-3
Ingram, D. (2002). The measurement of whole-word productions. *Journal of Child Language, 29,* 713–733. https://doi.org/10.1017/S0305000902005275
Kloots, H., Coussé, E., & Gillis, S. (2006). Vowel labelling in a pluricentric language: Flemish and Dutch labellers at work. In J. Van de Weijer & B. Los (Eds.), *Linguistics in the Netherlands* (pp. 126–136). Amsterdam, the Netherlands: Benjamins.
Kronenberger, W., Pisoni, D., Henning, S., & Colson, B. (2013). Executive functioning skills in long-term users of cochlear implants: A case control study. *Journal of Pediatric Psychology, 38,* 902–914. https://doi.org/10.1093/jpepsy/jst034
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 39,* 159–174. https://doi.org/10.2307/2529310
Leigh, J., Detmman, S., Dowell, R., & Briggs, R. (2013). Communication development in children who receive a cochlear implant by 12 months of age. *Otology & Neurotology, 34,* 443–450. https://doi.org/10.1097/MAO.0b013e3182814d2c
Leonard, L., Rowan, L., Morris, B., & Fey, M. (1982). Intra-word phonological variability in young children. *Journal of Child Language, 9,* 55–69. https://doi.org/10.1017/S0305000900003615
Macrae, T. (2013). Lexical and child-related factors in word variability and accuracy in infants. *Clinical Linguistics & Phonetics, 6–7,* 497–507. https://doi.org/10.3109/02699206.2012.752867
MacWhinney, B. (2000). *The childes project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.
McLeod, S., & Hewitt, S. (2008). Variability in the production of words containing consonant clusters by typical 2- and 3-year-old children. *Folia Phoniatrica et Logopaedica, 60,* 163–172. https://doi.org/10.1159/000127835
Mertens, P. (2001). *Fonilex*. Retrieved from http://bach.arts.kuleuven.be/fonilex/
Molemans, I. (2011). *A longitudinal investigation of aspects of the prelexical speech repertoire in young children acquiring Dutch: Normally hearing children and hearing-impaired children with a cochlear implant*. Antwerp, Belgium: University of Antwerp.
Moreno-Torres, I. (2014). The emergence of productive speech and language in Spanish-learning paediatric cochlear implant users. *Journal of Child Language, 41,* 575–599. https://doi.org/10.1017/S0305000913000056
Nikolopoulos, T., O'Donoghue, G., & Archbold, S. (1999). Age at implantation: Its importance in pediatric cochlear implantation. *The Laryngoscope, 109,* 595–599. https://doi.org/10.1097/00005537-199904000-00014
Nittrouer, S., Caldwell-Tarr, A., & Lowenstein, J. (2013). Working memory in children with cochlear implants: Problems are in storage, not processing. *International Journal of Pediatric Otorhinolaryngology, 77,* 1886–1898. https://doi.org/10.1016/j.ijporl.2013.09.001
Pisoni, D., Kronenberger, W., Roman, A., & Geers, A. (2010). Measures of digit span and verbal rehearsal speed in deaf children after more than 10 years of cochlear implantation. *Ear and Hearing, 32,* 60S–74S. https://doi.org/10.1097/AUD.0b013e3181ffd58e
R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/
Schauwers, K. (2006). *Early speech and language development in deaf children with a cochlear implant: A longitudinal investigation*. Belgium: University of Antwerp.
Schauwers, K., Taelman, H., Gillis, S., & Govaerts, P. (2008). Phonological proficiency and accuracy of young hearing-impaired children with a cochlear implant. In S. Kern, F. Gayraud, & E. Marsico (Eds.), *Emergence of linguistic abilities* (pp. 156–171). Newcastle, United Kingdom: Cambridge Scholars Publishing.
Sosa, A. (2015). Intraword variability in typical speech development. *American Journal of Speech-Language Pathology, 24,* 24–35. https://doi.org/10.1044/2014_AJSLP-13-0148
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
Sosa, A., & Stoel-Gammon, C. (2012). Lexical and phonological effects in early word production. *Journal of Speech, Language, and Hearing Research, 55,* 596–608. https://doi.org/10.1044/1092-4388(2011/10-0113)
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
Stoel-Gammon, C., & Sosa, A. (2007). Phonological development. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 238–256). Hoboken, NJ: Blackwell Publishing.
Taelman, H., & Gillis, S. (2002). Variation and consistency in children's truncation patterns. In J. Costa & M. J. Freitas (Eds.), *Proceedings of the gala 2001 conference on language acquisition* (pp. 263–270). Lisbon (Portugal): Associacào Portuguesa de Linguistica.
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*. Belgium: University of Antwerp.
Van Severen, L. (2012). *A large-scale longitudinal survey of consonant development in toddlers' spontaneous speech*. Belgium: University of Antwerp.
Verhoeven, J., Hide, O., De Maeyer, S., Gillis, S., & Gillis, S. (2016). Hearing impairment and vowel production : A comparison between normally hearing, hearing-aided and cochlear implanted Dutch children. *Journal of Communication Disorders, 59,* 24–39. https://doi.org/J.JCOMDIS.2015.10.007
Vihman, M., & McCune, L. (1994). When is a word a word? *Journal of Child Language, 21,* 517–542. https://doi.org/10.1017/S0305000900009442
Von Mentzer, C., Lyxell, B., Sahlén, B., Dahlström, O., Lindgren, M., Ors, M., . . . Uhlén, I. (2015). Segmental and suprasegmental properties in nonword repetition—An explorative study of the associations with nonword decoding in children with normal hearing and children with bilateral cochlear implants. *Clinical Linguistics & Phonetics, 29,* 216–235. https://doi.org/10.3109/02699206.2014.987926
Woltman, H., Feldstain, A., MacKay, C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology, 8,* 52–69.

Appendix A

Fixed Effect Results for the Variability Scores (Expressed in Logits) at Age 2;00

Fixed effects	Estimates and *SE*s	Consistent correct	Consistent incorrect	Variable with hits	Variable with no hits
Intercept	Estimate	–2.09**	–1.91***	–0.49**	–0.62***
	SE	0.27	0.18	0.17	0.23
Hearing status CI	Estimate	–0.32	0.83**	–0.81**	0.28
	SE	0.48	0.31	0.33	0.38
Syllable length	Estimate	–0.77***	–0.34	–0.93***	1.16***
	SE	0.23	0.18	0.15	0.13
Syllable length × Hearing status CI	Estimate	0.81	–0.26	0.82**	–0.70*
	SE	0.42	0.38	0.31	0.28

Note. Normal hearing is the reference category. *SE* = standard error; CI = cochlear implant.
*$p \le .05$. **$p \le .01$. ***$p \le .001$.

Appendix B

Fixed Effect Results for the Variability Scores (Expressed in Logits) at Age 3;00

Fixed effects	Estimates and *SE*s	Consistent correct	Consistent incorrect	Variable with hits	Variable with no hits
Intercept	Estimate	–0.72***	–2.19***	–0.18	–2.16***
	SE	0.21	0.24	0.16	0.27
Hearing status CI	Estimate	–0.78**	0.56	–0.46*	1.13**
	SE	0.28	0.29	0.21	0.35
Syllable length	Estimate	–0.11	0.23	–0.41*	0.64**
	SE	0.18	0.23	0.18	0.20
Syllable length × Hearing status CI	Estimate	–0.27	–0.35	–0.19	0.13
	SE	0.22	0.26	0.21	0.22

Note. Normal hearing is the reference category. *SE* = standard error; CI = cochlear implant.
*$p \le .05$. **$p \le .01$. ***$p \le .001$.

Appendix C

Fixed Effect Results for the Variability Scores (Expressed in Logits) at Age 4;00

Fixed effects	Estimates and *SE*s	Consistent correct	Consistent incorrect	Variable with hits	Variable with no hits
Intercept	Estimate	0.06	–2.30***	–0.60***	–3.17***
	SE	0.18	0.13	0.09	0.26
Hearing status CI	Estimate	–0.90**	0.49**	–0.02	1.64***
	SE	0.28	0.17	0.13	0.36
Syllable length	Estimate	–0.02	0.36*	–0.26*	0.33
	SE	0.10	0.15	0.12	0.20
Syllable length × Hearing status CI	Estimate	–0.39*	–0.17	–0.24	0.39
	SE	0.16	0.20	0.17	0.23

Note. Normal hearing is the reference category. *SE* = standard error; CI = cochlear implant.
*$p \le .05$. **$p \le .01$. ***$p \le .001$.

Appendix D

Fixed Effect Results for the Variability Scores (Expressed in Logits) at Age 5;00

Fixed effects	Estimates and *SE*s	Consistent correct	Consistent incorrect	Variable with hits	Variable with no hits
Intercept	Estimate	0.08	–2.29***	–0.70	–3.56***
	SE	0.27	0.16	0.47	0.50
Hearing status CI	Estimate	2.11***	–0.38	–4.54***	–2.29*
	SE	0.4	0.26	0.99	1.08
Syllable length	Estimate	0.16	0.17	–0.43**	0.51*
	SE	0.13	0.19	0.15	0.25
Syllable length × Hearing status CI	Estimate	–0.21	–0.06	0.40	–0.70
	SE	0.23	0.28	0.39	0.62

Note. Normal hearing is the reference category. *SE* = standard error; CI = cochlear implant.
*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.