Accepted Manuscript

Title: Phonemic accuracy development in children with cochlear implants up to five years of age by using Levenshtein Distance

Author: Jolien Faes Joris Gillis Steven Gillis

PII: S0021-9924(15)30007-1
DOI: http://dx.doi.org/doi:10.1016/j.jcomdis.2015.09.004
Reference: JCD 5732

To appear in: *JCD*

Received date: 6-1-2015
Revised date: 18-9-2015
Accepted date: 30-9-2015

Please cite this article as: Faes, J., Gillis, J., and Gillis, S.,Phonemic accuracy development in children with cochlear implants up to five years of age by using Levenshtein Distance, *Journal of Communication Disorders* (2015), http://dx.doi.org/10.1016/j.jcomdis.2015.09.004

Title

Phonemic Accuracy Development In Children With Cochlear Implants Up To Five Years Of Age By Using Levenshtein Distance

Running Title

Phonemic Accuracy Of Ci Children: Levenshtein Distance

Authors

Jolien Faes

Computational linguistics & psycholinguistics (clips) research centre, university of antwerp, Belgium

e-mail: jolien.faes@uantwerpen.be

Joris gillis

Mobile vikings belgium, hasselt, belgium

E-mail: joris.gillis@gmail.com

Steven gillis

Computational linguistics & psycholinguistics (clips) research centre, university of antwerp, Belgium

E-mail: steven.gillis@uantwerpen.be

Corresponding author

Steven gillis

Prinsstraat 13

B-2000 antwerpen, belgium

Phone: +32(0)3 265 5240

E-mail: steven.gillis@uantwerpen.be

Abstract

Phonemic accuracy of children with cochlear implants (CI) is often reported to be lower in comparison with normally hearing (NH) age-matched children. In this study, we compare phonemic accuracy development in the spontaneous speech of Dutch-speaking children with CI and NH age-matched peers. A dynamic cost model of Levenshtein distance is used to compute the accuracy of each word token. We set up a longitudinal design with monthly data for comparisons up to age two and a cross-sectional design with yearly data between three and five years of age. The main finding is that phonemic accuracy steadily increases throughout the period studied. Children with CI's accuracy is lower than that of their NH age mates, but this difference is not statistically significant in the earliest stages of lexical development. But accuracy of children with CI initially improves significantly less steeply than that of NH peers. Furthermore, the number of syllables in the target word and target word's complexity influence children's accuracy, as longer and more complex target words are less accurately produced. Up to age four, children with CI are significantly less accurate than NH children with increasing word length and word complexity. This difference has disappeared at age five. Finally, hearing age is shown to influence accuracy development of children with CI, while age of implant activation is not.

Keywords: Phonemic accuracy; children with CI; Levenshtein distance; target word complexity; target word syllable length

Introduction

This study describes the development of phonemic accuracy of children with a cochlear implant (henceforth, CI) acquiring Dutch as their native language. As speech production depends on speech perception (Altvater-Mackensen & Fikkert, 2010; Jusczyk, 1992; Stoel-Gammon, 2011; Stoel-Gammon & Sosa, 2007), spoken language development in children with CI is a topic that has received considerable attention in the literature. The ultimate question is whether the spoken first language proficiency of children with CI eventually reaches a level comparable to that of normally hearing (henceforth, NH) age mates. In this respect Nicholas and Geers (2007) published an aptly titled paper: “Will they catch up?” in which they investigated whether children with a CI, after an initial delay in comparison to NH children, eventually close the gap and achieve age appropriate spoken language levels in terms of vocabulary, sentence complexity and morphology. In this paper, we elaborate on this question by investigating speech production of children with CI and more specifically phonemic accuracy. Phonemic accuracy is based on phonemic transcriptions instead of narrow phonetic transcriptions. In other words, we focus on broad production accuracy of phonemes, regardless of phonetic variance. Speech production accuracy is important since it affects children’s speech intelligibility (Ingram, 2002) and, consequently, their communicative effectiveness. The development of phonemic accuracy is traced in a group of children with CI and a control group of NH children, using Levenshtein distance, a measure that is relatively new in language acquisition research. We study three research questions: (1) How does phonemic accuracy develop immediately after cochlear implantation up to age 2;0? (2) How does it develop with prolonged linguistic experience, i.e., up to five years of age? And (3) Are there effects of length of implant use and age at implantation in accuracy development of children with CI. In what follows, we first discuss various measures of phonemic accuracy used in the literature and then, we elaborate on the main findings of phonemic accuracy development in the literature.

Measures of phonemic accuracy

In the literature two methods are frequently used to express phonemic accuracy: *Percentage of Consonants Correct* (PCC, Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997; Shriberg & Kwiatkowski, 1982) and *phonological mean length of utterance* (pMLU, Ingram, 2002). Both measures increase when production accuracy improves (Tyler & Lewis, 2005).

Shriberg and Kwiatkowski (1982) developed the first phonological yardstick, PCC. This metric indicates the percentage of correctly produced consonants. The procedure is fairly simple: A child’s rendition of a particular adult target word is lined up with a transcription of the actual target, and the number of overlapping consonants is counted. For instance, if a child utters /b□k/ for the adult target Dutch word /b□ks/ ‘fight’, two out of three consonants are correctly rendered. This procedure is repeated for all the words in a speech sample and the end result is the percentage of correctly produced consonants.

Although PCC is frequently used in child language studies, this metric faces three problems. First of all, only consonants are taken into consideration, while vowels are not considered, which means that the measure only partly reflects the child's accuracy. This problem can easily be dealt with by the *Percentage of Phonemes Correct* (PPC, Shriberg et al., 1997), which considers consonants as well as vowels. Secondly, in both PCC and PPC, a rendition of a particular target word may differ from the intended target in various ways. It may involve deletions, insertions and/or substitutions. In Table 1, the target word /b□ks/ 'fight' is compared with (example) renditions involving a deletion, a substitution and two examples of insertions. Deleting one consonant (as in /b□ks/) or substituting one (as in /p□ks/) leads to a PCC accuracy score of 66%. This result is expected in the sense that both renditions do not equal the adult target, and hence an accuracy score of 100% should not be expected. Inserting a consonant or a vowel (as in /b□kəs/ and /br□ks/) leads to a higher PCC accuracy score. In fact, in the case of solely insertions, the PCC is 100%, which implies that the child's rendition equals the adult target, but it does not. So the second problem with PCC is that substitutions and deletions are "penalized" as errors or deviations from the adult target, while insertions are not. In fact, this means that implicitly there is a weighting of the errors or deviations. Some errors have a more heavy weight than others. In the case of PCC insertions do not influence the value of PCC at all, while substitutions and deletions have an equal influence on the value of PCC. Nevertheless, such errors have varying effects on speech accuracy and speech intelligibility. Thirdly, the phonemic accuracy of less complex words, such as monosyllabic words or words with singleton consonants, is expected to be higher than the accuracy of more complex words, such as multisyllabic words or words with consonant clusters. For instance, Gathercole, Willis, Emslie, and Baddeley (1991) pointed out that accuracy decreases with increasing target syllable length. Including word length may have an added value, but in computing PPC word length is not taken into account.

***Table 1*. Examples of PCC calculations**

			PCC child	*PCC*	*Accuracy rate*
Target		/b□ks/ *fight*		3	
Replica	*1) Deletion*	/b□k/	2		66%
	2) Substitution	/p□ks/	2		66%
	3) Insertion	/b□kəs/	3		100%
		/br□ks/	3		100%

***Table 2.* Examples of pMLU calculations**

			pMLU child	*pMLU*	*PWP*
Target		/b□ks/ *fight*		7	
Replica	*1) Deletion*	/b□k/	5		0.71 (71%)

2) Substitution	/p□ks/	6	0.86 (86%)
3) Insertion	/b□kəs/	7	1.00 (100%)

In 2002, Ingram proposed phonological mean length of utterance (pMLU) as an alternative phonological measure. pMLU combines word length with PCC. pMLU is calculated by counting the number of phonemes of each word production, regardless of their accuracy. This means for instance that if the adult target contains three segments (/k□t/, 'cat') and the child produces three segments then the child's rendition is credited with three pMLU points. An additional point is added to the pMLU score for each correctly produced consonant (Ingram, 2002). To measure the degree of accuracy, Ingram (2002) developed the proportion of whole-word proximity (PWP). PWP is calculated by dividing the child's pMLU by the target language's pMLU. If the child's rendition equals the adult target, PWP equals 1: What is targeted and what is actually produced are identical. Therefore, PWP is "an indirect measure of the child's intelligibility" (Ingram, 2002, p. 718). In Table 2, the pMLU and PWP scores of three child renditions of the target word /b□ks/ are given as an example. Both deleting and substituting a consonant lead to an imperfect PWP score (lower than 1.00). However, the insertion of /ə/ in /b□kəs/, as shown in example (3), is not penalized, as the child's pMLU score equals the target pMLU. Consequently, a maximum PWP score of 1.00 is reached.

pMLU faces three problems. First of all, pMLU reflects not only phonological but also morphological development (Taelman, Durieux, & Gillis, 2005). At least for Dutch, pMLU is higher when inflected words, like nouns plurals, finite verb forms etc., are included in the corpus (Taelman et al., 2005). Secondly, since pMLU is dependent on word length, languages with inherently longer words will have higher pMLU rates. This was illustrated by Saaristo-Helin, Savinainen-Makkonen, and Kunnari (2006), who report higher pMLU rates for Finnish than English, with the former having longer words than the latter. This does not only make cross-linguistic comparison of children's pMLU highly problematic, it also has the pernicious effect of the same error being weighted differently according to language. If a child deletes a phoneme in a language with longer words, the pMLU and associated PWP will be higher than if a deletion had occurred in a language with shorter words. Therefore, the same error incurs a higher penalty in the second language. Thirdly, as for PCC, an implicit weighting of speech errors is present, as Table 2 indicates. While deletions and substitutions are penalized as errors, insertions are not. The insertion of /ə/ in /b□kəs/ is not penalized and has no influence on the accuracy rate (100%). Moreover, the pMLU metric penalizes deletions more heavily than substitutions. Nevertheless, a more explicit and desirable weighting in terms of frequency and complexity of errors is absent. The same is true for PCC/PPC. Speech errors differ from one another both in frequency and severity. Firstly, some errors are more common in children than other ones. Secondly, some speech errors are more severe as they influence intelligibility more than other speech errors. Therefore, it might be advisable to penalize frequent and less severe errors less than infrequent and severe errors. These are integrated into the Levenshtein distance measure.

In this paper, Levenshtein distance (LD) is used as an alternative to track production accuracy and its development. LD is a commonly used technique to measure the difference between character strings in various scientific fields such as computer science, bio-informatics, and dialectology. LD between two character strings expresses their difference. A small distance between two strings indicates that they are very similar, and a large distance means that they are dissimilar. In the present context LD is used to compute the distance between the phonemic transcription of an adult target word and the phonemic transcription of a child's rendition of that target word. It can be expected that the distance between the adult target and the child's rendition is relatively large when a child starts uttering conventional words. And over time, the distance is expected to diminish, meaning that the child's productions become more and more adult like. Without going into the technical details of LD (see Method section), LD is determined by counting the number of edit operations needed to transform one string of characters into another one (Heeringa, 2004; Nerbonne & Heeringa, 2010; Wieling, Margaretha, & Nerbonne, 2011). Three edit operations are assumed: Substitutions, deletions and insertions. The more edit operations that are needed to transform one string (i.c. an adult target word) into another (i.c. the child's rendition of that word), the more distant these two strings are. Moreover each edit operation has a cost associated with it and the number of edit operations needed determines the LD between the two character strings. For example, suppose the cost of each edit operation is 1, consider the difference between /b□t/ and /b□/. The difference between these two character strings is the deletion of /t/, so LD equals 1, because the other characters are identical.

LD can not only be used to measure the distance between two character strings, DNA sequences or dialectal variants, but also to measure the distance between a word spoken by a child and the adult pronunciation of that word. Only a few studies used LD to measure phonemic accuracy. Riches, Loucas, Baird, Charman, and Simonoff (2011) showed that LD error rates of adolescents with Specific Language Impairment are higher than adolescent with Autism plus Language Impairments. Sanders and Chin (2009) showed that the distance between word productions of children with CI and the adult targets measured by LD correlates with intelligibility judgments of naïve listeners. The larger the distance the less intelligible the children's speech was judged. In this paper phonemic accuracy, as measured by LD, is studied in a group of children with CI and a group of NH children. The use of this measure is relatively new in studies on language acquisition and has some important advantages. First of all, insertions are taken into account in the LD measure, in contrast to PCC and pMLU. In other words, all edit operations are considered in LD, while this is not the case in PCC and pMLU. Moreover, in the current paper, a dynamic algorithm of LD is used. This means that the LD computations are based on a model of adult spoken language and not on a priori defined weights. In adult language, some phonemes typically appear more frequently than others, some phonemic variations are frequent in spontaneous speech (e.g., the deletion of /n/ after schwa at the end of words, as in /etə/ instead of /etən/ 'eat'). These frequency differences are taken into account when computing the distance between the adult model and the child's rendition. Infrequent speech errors as well as severe speech errors are penalized more heavily than frequent and less severe ones, due to the algorithm's way of working (this is fully explained in the method section (2.3.).

Phonemic accuracy in NH and children with CI

In this paper, we study the phonemic accuracy of children with CI immediately after implantation up to the age of 2;0 (years;months) and with prolonged linguistic experience (i.e. up to age 5;0) and compare these to NH peers. In NH children phonemic accuracy reaches ceiling percentages around the age of 3;0 (English: Dodd, Holm, Hua, & Crosbie, 2003; Irwin & Wong, 1983; Finnish: Saaristo-Helin, 2009). For English, Warner-Czyz (2005) has shown that production accuracy in spontaneous speech is higher in NH children as compared to children with CI up to six months after meaningful word onset. In spontaneous speech, the accuracy of Dutch speaking NH children is found to be higher in comparison to children with CI. Between ages 2;1 and 2;6 pMLU and PWP reach median scores that are significantly higher in NH children (pMLU: Ca. 6.3, PWP: Ca. 0.8) than in children with CI (pMLU: Ca. 4.5, PWP: Ca. 0.7) (Schauwers, Taelman, Gillis, & Govaerts, 2008). At age 2;0, median proportions of phonemic accuracy at the word level are significantly higher for NH children (0.66) than for age-matched children with CI (0.58) (Van den Berg, 2012). Age at implantation of the participants with CI reported in Schauwers, Taelman, et al. (2008) and in Van den Berg (2012) was 1;0 (SD = 0;5). In the present article, the same children with CI as in Schauwers, Taelman, et al. (2008) and Van den Berg (2012) were studied. We expand their research in two ways. Firstly, the children with CI are studied up age 5;0, and secondly, a more fine-grained accuracy measure, viz. Levenshtein distance, is used. Therefore we expect the children with CI to be less accurate than NH peers at age 2;0.

Nicholas and Geers (2007) report that English speaking children with CI implanted at around age 2;6 are more likely to catch up with age-matched NH peers at age 4;6 than later implanted children with CI. Their conclusion holds for several linguistic domains, namely morphology, syntax and vocabulary. In the present article, we investigate whether Nicholas & Geers' finding can be replicated in another linguistic domain, viz. the phonemic accuracy development of early implanted children with CI. As children grow older their renditions of words can be expected to approximate the adult target forms. We investigate whether in the age span studied, children with CI reach a comparable level of accuracy as NH children.

Studies of phonemic accuracy of children with CI acquiring Dutch as compared to their NH peers are restricted to very young children (under age 2;0 in Schauwers, Taelman, et al. (2008), Van den Berg (2012)). In children acquiring English, production accuracy at older ages is shown to be significantly higher in NH children than in children with CI. However, these results should be considered with care when generalizing them to Dutch. Approximately at age 3;6, PPC is 83% for NH children and 53% for children with CI (Ertmer, Kloiber, Jung, Kirleis, & Bradford, 2012). At age 4;0, the same trend emerges in the PCC of word initial consonants in a short sentence repetition task. For NH children, all initial consonants except fricatives and affricates (86% accuracy) reach ceiling accuracy, while for children with CI initial consonant accuracy is only 62% (Ertmer & Goffman, 2011). At ages 3;0, 4;0 and 5;0, a Goldman-Fristoe Words and Sounds Test of Articulation (GFTA-2 Words and Sounds) score of 108.05 in NH children and only 90 in hearing impaired children is reported by Eriks-Brophy, Gibson, and Tucker (2013). This difference is statistically significant. However, 72% of the hearing impaired children reached average scores at age 5;0 (Eriks-Brophy et al., 2013) and thus caught up with their NH peers. Note, however, that in Eriks-Brophy et al. (2013) and in Ertmer et al. (2012) mean age at implantation was 2;0 (SD = 1;1) and 1;6 (SD = 0;6) respectively, whereas in our study, children with CI were implanted much earlier, resulting in a mean age at implantation of 1;0 (SD = 0;5). Moreover, the hearing impaired participants reported on in Eriks-Brophy et al. (2013) were children with cochlear implants as well as children with hearing aids. All hearing impaired children were compared as a group to NH children. In contrast, we compare NH children with children with CI only. Moreover, we follow Dutch-speaking children with CI and NH children up to the age of five. Unlike Ertmer and Goffman (2011), Ertmer et al. (2012) and Eriks-Brophy et al. (2013), who use standardized tests, we study spontaneous speech of children with CI and NH children to compare accuracy development.

Numerous factors affect phonemic accuracy, including target word length and target word complexity. The number of syllables in the target word as well as the complexity of the target, determined by i.a. the presence of consonant clusters, influence children's accuracy. This influence is attested in various studies reported in literature using nonword repetition tasks. For instance Gathercole et al. (1991) reported decreasing phonemic accuracy in typically developing children with increasing target nonword syllable length. In a repetition task, they were less accurate when target nonword syllable length increased from two to four syllables. Burkholder-Juhasz, Levi, Dillon, and Pisoni (2007) and Von Mentzer et al. (2015) found similar results for children with CI and Nittrouer, Caldwell-Tarr, Sansom, Twersky, and Lowensthein (2014) found similar results for both NH children and children with CI. Carter, Dillon, and Pisoni (2002) indicated that syllable length in the target nonword influences accurate repetition of suprasegmental features in a nonword repetition task in children with CI. The reproduction of the correct number of syllables and the reproduction of the correct primary stress decreases when the target nonword has more syllables (Carter et al., 2002). Next, while Carter et al. (2002) did not find an effect of nonword complexity on suprasegmental accuracy in children with CI, nonword complexity influences phonemic, and thus segmental, accuracy in typically developing children (Macrae, 2013) and in children with CI (Von Mentzer et al., 2015). Macrae (2013) found lower consonant accuracy scores in nonwords with late developing sounds like fricatives, liquids and affricates and nonwords with consonant clusters; Von Mentzer et al. (2015) found more consonant omissions and substitutions when consonant clusters were present in the target nonword. Thus, in NH children and children with CI, consonant accuracy decreases with increasing phonological complexity. Furthermore, frequent phonological processes in child language emphasize the influence of target word length and target word complexity on accuracy. For instance, reduplication, weak syllable deletion and cluster reduction occur frequently in child language (Johnson & Reimers, 2010): Monosyllabic, but mainly multisyllabic words are simplified by reduplicating the first syllable or by deleting weak unstressed syllables. In words with consonant clusters, clusters are simplified and only one consonant usually remains (Johnson & Reimers, 2010; Jongstra, 2003). Due to these phonological processes, children's accuracy decreases considerably. The effects of target word complexity and target word syllable length were shown in nonword repetitions tasks for both NH children and children with CI. In the present paper, however, the influence of both factors on the phonemic accuracy of NH children and children with CI's spontaneous speech is examined.

For children with CI, two factors have been shown to affect language development, namely age of implantation and length of implant use, i.e. 'hearing experience'. Numerous studies have pointed out the benefit of early implantation on language development (e.g. Nicholas & Geers, 2007; Schauwers, Gillis, & Govaerts, 2008), but only a few studies examined its effect on phonemic accuracy (Connor, Craig, Raudenbush, Heavner, & Zwolan, 2006; Schauwers, Taelman, et al., 2008; Van den Berg, 2012). For Dutch, Van den Berg (2012) found lower accuracy scores in later implanted children (range 5 – 20 months) and Schauwers, Taelman, et al. (2008) found lower pMLU scores in children implanted in the second year of life as compared to children implanted in the first year of life. For English, Connor et al. (2006) concluded that children implanted before age 2;6 have better consonant production accuracy than later implanted age-mates. Though age of implantation is a major factor, length of implant use has also been indicated to influence language development (e.g. Szagun & Stumper, 2012). Szagun and Stumper (2012) even report that age of implantation does not affect language development of children with CI implanted within the sensitive period, i.e. before age 4;0, in contrast to length of implant use. For English-speaking children, accuracy is reported to increase with longer implant use (Blamey et al., 2001; Eriks-Brophy et al., 2013; Gantz, Tyler, Woodworth, Tye-Murray, & Fryauf-Bertschy, 1994; Tobey, Geers, Brenner, Altuna, & Gabbert, 2003; Tomblin, Spencer, & Lu, 2008; Tye-Murray, Spencer, & Woodworth, 1995). After four years of implant use, overall phonemic accuracy is 62.9% (Tomblin et al., 2008). Accuracy increases after six years of implant use to 76.28% according to Tomblin et al. (2008) and approximately 86% according to Blamey et al. (2001). In Blamey et al. (2001), mean age at implantation was 3;9 (SD = 0;11) and in Tomblin et al. (2008), mean age at implantation was 4;6 (SD = 2;1). In the present paper, we compare long-term accuracy development of NH children and children with CI up to age 5;0, and at that age the children with CI have up to 4;6 years of device use. To the best of our knowledge, long-term accuracy development of Dutch children with CI has not been studied yet.

The present paper studies the impact of implantation in the first two years of life after early detection. As mean age at implantation in Blamey et al. (2001) and Tomblin et al. (2008) were above three years of age, the researchers were unable to investigate the influence of early implantation. In our study, phonemic accuracy of children with CI implanted at a much younger age (before age 1;8) is investigated, allowing us to study the impact of early implantation on accuracy. We expect the children with CI to benefit from their early implantation, eventually resulting in children with CI catching up with their NH peers by age 5;0. Additionally, the impact of length of implant use on accuracy is examined.

Method

Participants

Nine children with CI were studied longitudinally from the moment their device was activated (median = 1;0, range 0;6 – 1;9) up to age 5;0. The children were followed monthly from the moment of implant activation up to 2;6 years after activation. Additionally, yearly data were collected between ages 3;0 and 5;0. Here, the data will be analyzed from the appearance of first spoken words (median = 1;6, range 1;3 – 1;11).

All children were monolingual Dutch and 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 before implantation. No other patent health or developmental problems were reported at the time of testing. All participants received a Nucleus-24 implant before age 1;8. The mean implant age was 1;0 (SD = 0;5) and the mean age of CI activation was 1;1 (SD = 0;5). After implantation, the mean PTA improved to 32.33 dB HL (SD = 7.11) at age 5;0. All children used oral communication with only a limited amount of sign language. Six out of 9 children with CI received a second CI within the period studied. Detailed information can be found in Table 3.

A control group of 30 NH children was followed monthly and longitudinally between ages 0;6 and 2;0 as part of a larger study of speech and language development in the first two years of life. Here, the data will be examined from the appearance of the first spoken words (median = 1;2, range 1;2 – 1;5). In addition, a cross-sectional design was set-up for NH children between ages 3;0 and 5;0. A total of 32 NH children participated: 9 three year olds (mean = 3;0, SD = 0;1), 12 four year olds (mean = 4;0 months, SD = 0;1), 11 five year olds (mean = 5;0, SD = 0;1). Thus, these 3-, 4- and 5-year-old children were recorded only once at home, interacting with (predominantly) their mothers. The participants were all native speakers of Dutch. None of the NH children had patent hearing problems, as checked with an auto-acoustic emission test by *Kind&Gezin*, the Flemish infant welfare center. This hearing screening took place approximately three weeks after birth, as part of a nation wide neonatal screening after birth (Desloovere, Verhaert, Van Kerschaver, & Debruyne, 2013).

Table 3. Characteristics of the children with CI

ID	**Gen-**	**Cause of**	**PTA**	**PTA**	**Age 1st**	**Age**	**Age 2nd**	**Age**	**Length**

	der	deafness	unaided	CI	CI	activation 1e CI	CI	first word	implant use (Age 5;0)
S1	F	MCG 26	120	35	1;1	1;3	6;3	1;8	3;9
S2	F	MCG 26	120	27	0;6	0;8	4;8	1;4	4;4
S3	F	CMV	115	25	0;10	1;0	5;10	1;8	4;0
S4	M	MCG 26	113	42	1;6	1;7	-	1;8	3;5
S5	M	MCG 26	93	32	1;5	1;6	6;4	1;6	3;6
S6	M	MCG 26	120	37	0;9	0;10	-	1;4	4;2
S7	F	Genetic	117	23	0;5	0;6	1;3	1;3	4;6
S8	F	Unknown	112	42	1;7	1;9	-	1;11	3;3
S9	F	MCG 31	103	28	0;9	0;10	1;11	1;3	4;2
Mean			*112.56*	*32.33*	*1;0*	*1;1*	*4;4*	*1;6*	*3;11*
SD			*9.12*	*7.11*	*0;5*	*0;5*	*2;3*	*0;3*	*0;5*

MCG = mutation in connexion-gene, CMV = Cytomegalovirus

PTA = Pure Tone Average (in dB HL = decibel hearing level)

Ages are represented in years;months

- = no second CI

Procedure

The data used in the current paper are gathered as a part of a larger corpus, namely the CLiPS Child Language Corpus (CCLC). This corpus consists of on average one-hour video recordings of spontaneous, unstructured parent-child interactions at the children's homes. These video recordings started at age 0;6 for the NH children and from the month of implant activation for the children with CI. Video recordings were made on a monthly basis up to age 2;0 for the NH children and up to 2;6 years after implantation for the children with CI. Additional video recordings were made at ages 3;0, 4;0 and 5;0 for both children with CI and (other) NH children. In the present paper, only samples in which lexical items appeared were analyzed. In other words, prelexical vocalizations were not analyzed in the current paper.

Approximately 20 minutes of each complete recording was selected for transcription and analyses. Immediately after a recording session the episodes in which the child was most voluble were selected. This data reduction step was taken in order to keep the transcription time within reasonable limits. On average, the transcription process lasted 14 hours, from the video recording at the child's home up to the complete transcription of each 20 minute selection (Molemans, Van den Berg, Van Severen, & Gillis, 2012; Schauwers, 2006; Van den Berg, 2012; Van Severen et al., 2013). Those 20 minute selections were transcribed in the software package CLAN according to the CHAT conventions (MacWhinney, 2000). For the identification of words, the procedure articulated by Vihman and McCune (1994) was used. More specifically, in order to be counted as a word, a child's production has 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 comprise for instance an exact match of the vocalization with the target form, or a child's form that deviates from the target following one of the frequent phonological processes attested in Dutch (Beers, 1995; Gillis & Schaerlaekens, 2000). Next, the criteria based on context involve for instance maternal identification, i.e. the identification of the vocalization as a word by the mother judging from her reply to the child. Finally, the criteria based on the relation to other vocalizations cover for instance the absence of inappropriate use, i.e. the vocalization is only used in plausible contexts.

For the transcription of the children's productions, both an orthographic and a phonemic transcription was made based on the video recordings. A phonemic transcription of the target words was added automatically to the children's productions from the lexical database Fonilex, which is 'a pronunciation database containing the phonetic transcription of the most frequent word forms of Dutch as spoken in Flanders' (Mertens, 2001). After target words were added, the child's utterances were aligned with the target forms at the phoneme level. The phonetic characters of the transcription pairs were automatically aligned with a computer program implementing a dynamic alignment algorithm (based on ADAPT (Elffers, Van Bael, & Strik, 2005)). The alignments were verified manually and corrected if needed. For the transcription of the adult's productions, an orthographic transcription was made based on the video recordings. The phonemic transcription was retrieved and added automatically from the Fonilex database (Mertens, 2001). This phonemic transcription was verified manually and specific attention was given to phenomena of spontaneous Dutch such as the deletion of final /n/ in Dutch. For the adult transcriptions, target words were also added and aligned with the adult's production, using the exact same procedure as for the children.

For the children with CI, a total of 58,686 word tokens, with a median of 5,606 per child (range: 4,079 – 10,520) were available. Up to age 2;0, 3,406 word tokens were available, with a median of 320 word tokens per child (range 16 – 1140). Between ages 3;0 and 5;0, a total of 29,326 word tokens were available, with a median of 2,944 per child (range 1,927 – 5,002). For the longitudinal analysis of only children with CI, all 58,686 word tokens were used. For the NH children, a total of 59,019 word tokens were available. For the NH children up to age 2;0, data consisted of a total of 42,535 word tokens, with a median of 1438.5 word tokens per child (range 455 – 2,889). For the NH children between ages 3;0 and 5;0 a total of 16,484 word tokens were available, with a median of 525.5 word tokens per child (range 178 – 965). Reliability of phonemic transcriptions was computed on 10% of the corpus. With respect to percentage of agreement, interrater reliability was 63.69% and intrarater reliability 81.51% for NH speech samples. For the corpus of children with CI, only interrater reliability was checked and equals 81.63%. In line with Cucchiarini (1996) also Kappa scores were calculated in order to consider the possible influence of chance. Kappa scores were 0.60 for interrater reliability in the NH speech samples and 0.80 for intrarater reliability in the NH speech samples. These scores are on the edge of "moderate" to "substantial" and on the edge of "substantial" to "almost perfect" respectively (Landis & Koch, 1977). The Kappa score for intrarater reliability of the CI speech samples equals 0.87 and is interpreted as "almost perfect" (Landis & Koch, 1977). Further and more detailed information on participants, data collection and reduction and data transcription can be found in Molemans et al. (2012), Schauwers (2006) and Van den Berg (2012).

Levenshtein distance (LD)

Conceptually, computing Levenshtein distance amounts to comparing word after word a phonemic transcription of a child's own word productions with the adult equivalent of those words. If both transcriptions are identical, then their distance is in principle zero and the child's production can be said to be adult-like. If that is not the case, LD measures the distance between the adult target and the child's rendition of that target. LD is a way to measure that distance by computing the minimal edit distance between the two transcriptions: How can one transcription be transformed into the other by a (minimal) set of edit operations (deletions, insertions, substitutions)? Given that set of operations, each operation is given a weight or "cost". Wieling, Prokic, and Nerbonne (2009) proposed a procedure for inducing those weights automatically from a corpus of transcribed speech, instead of (arbitrarily) assigning a weight to each edit operation a priori. Their proposal is essentially the following: Suppose a target word contains /e/ and that segment is rendered as /ə/, then that rendition is closer to the target than when /e/ would have been substituted by /u/. For instance, in Dutch vowel reductions – such as rendering /e/ as /ə/ – are frequent in spontaneous adult speech (Swerts, Kloots, Gillis, & De Schutter, 2003). Hence such substitutions are expected to occur frequently when comparing a standard transcription deriving from a phonemic lexicon with a (broad) transcription of spontaneous speech, while the substitution of /e/ by /u/ is very infrequent. Hence the distance between /e/ and /ə/ is smaller than that between /e/ and /u/. In other words, the "cost" for transforming /e/ into /ə/ should be smaller than the "cost" for transforming /e/ into /u/. This cost model is dynamically derived from the corpus, and frequency information is crucial in computing the model: The "cost" for transforming a segment into another one is smaller if this pairing occurs relatively frequently. Note in bypassing that Wieling, Margaretha, and Nerbonne (2012) found strong significant correlations between the induced distances (or costs) and the acoustic distances they measured.

How is LD computed in the present study? The basic algorithm is taken from Wieling et al. (2012); Wieling et al. (2009): LD is computed using a dynamic cost model. In what follows the algorithm will be described and exemplified first. After the description of the basic algorithm, the adjustments that were made to fit the purposes of the current study will be dwelled upon. Computing the distance between an adult target and a child's rendition of that target, is not exactly the same as computing the distance between two adult dialects as was the case for Wieling et al. (2009). Moreover, the corpus analyzed in the present study is a longitudinal one, which requires some additional measures to be taken.

The first phase in computing LD consists of aligning two strings of segments, such as an adult target word and a child's rendition of that target. For instance, the pair of transcriptions in (1) represent the target word /spelə/ (the Dutch word for *to play*) and a child's rendition of that word as /pe/:

(1) Adult target s p e l ə
Child's rendition p e

In principle there are many different possible alignments, but the algorithm incorporates a binary same-different strategy, trying to line up matches between segments and avoiding mismatches, and trying to maximize the number of matches. Moreover, the algorithm uses a VC-sensitive strategy allowing only vowels to be lined up with vowels and consonants with consonants (Wieling et al., 2009). Practically speaking the algorithm is implemented using a dynamic programming algorithm (Wagner & Fisher, 1974) that seeks the minimal edit distance. Applying these principles to example (1), the net result will be (2).

(2) Adult target s p e l ə
Child's rendition . p e . .

Note that both the target and the rendition in (2) are of equal length since "filler characters" (represented by a dot) have been inserted in the shorter character string, which stand for an empty position or "zero segment". The /p/ and /e/ are aligned with the matching segments in the target and the rendition, /s/ in the target has been deleted in the child's rendition (introducing a dot), /l/ and /ə/ are deleted as well as.

The second step consists of determining the cost of an insertion, substitution or deletion. This is accomplished by implementing a dynamic cost model (Wieling et al., 2012; Wieling et al., 2009). For this purpose the program runs through a transcription and considers each target – rendition pair and computes the Pointwise Mutual Information (PMI) of each segment. For instance, given the initial alignment represented in (2), the probability of the pairings (/s/ - /./), (/p/ - /p/), (/e/ - /e/), etc. are computed. The PMI of all these pairs is calculated using equation (3):

(3)
$$PMI(x,y) = log_2 \frac{p(x,y)}{p(x)p(y)}$$

where p(x,y) denotes the probability of encountering the pair (x,y) in the alignment, p(x) the probability of encountering segment /x/, and p(y) the probability of segment /y/. Take the pair (/s/ - /./) in example (2), the PMI of that pair is the probability of /s/ in the target being lined up with the empty character /./ in the child's rendition, divided by the product of the probability of /s/ in the adult corpus and the probability of /./ in the child corpus. Mutatis mutandis, the PMI of the pair (/p/ - /p/) is the probability of /p/ in the target paired with /p/ in the rendition, divided by the probability of /p/ in the adult corpus and its probability in the child corpus. The division is actually meant to normalize the probability of the pairing p(x,y) in (3), or the statistical dependence of segments /x/ and /y/, with respect to the probability of /x/ and /y/ being statistically independent (Wieling et al., 2012).

This basic procedure is repeated for the consecutive child utterances and their targets in a transcription. Evidently, the PMI value for each pair of segments will change as more words are processed. For instance, in addition to the pairing (/p/ - /p/) in (2), alternative pairings may be encountered such as (/p/ - /b/) or (/p/ - /./), i.e., we may encounter examples of /p/ being substituted by /b/, or /p/ being deleted. After processing an entire transcription, the net result is a first alignment of both transcriptions, and a list of segment pairs, each with a PMI value. The PMI value of a pair is converted into a cost (or a distance, for that matter) by subtracting it from the maximum PMI value. The rationale is that if the PMI value is very high, meaning that a particular pairing of segments occurs very frequently in the data, the resulting cost (distance) should be small. In this way, frequently co-occurring segments will have a much smaller cost (distance) than segments that do not co-occur very often.

As an example of how LD is computed, consider the two child renditions of the adult word /spelə/ (Eng.: *to play)* in (4):

(4) Adult target s p e l ə

Child's rendition 1 . p e . .

Child's rendition 2 . p e l ə

In order to calculate the LD of the 2 child renditions (/pe/ and /pelə/), the proportions of each segment and each pair of segments in the alignment have to be computed first. Take the pair /s/-/./: suppose the proportion of /s/ in the transcription of the adult targets equals 0.075 and the proportion of /./ in the child's renditions equals 0.05, and in the aligned transcriptions the proportion of the pair /s/ - /./ equals 0.004, meaning that in 0,4% of the cases the child deletes /s/. Inserting these figures in equation (3) results in the PMI: $\log_2 (0.004 / 0.075 * 0.05) = 0.09$. By way of example, proportions are given for the relevant segments and pairs in (4), and the PMI of each pair is calculated using the formula in (3). The result is shown in (5):

(5)

	Pair	P(x,y)	P(x)	P(y)	PMI
a.	/s/ - /./	0.0040	0.0750	0.0500	0.09
b.	/p/ - /p/	0.9900	0.0005	0.0005	21.92

c.	/e/ - /e/	0.3500	0.2500	0.2500	2.49
d.	/l/ - /l/	0.3000	0.0900	0.0900	5.21
e.	/l/ - /./	0.0100	0.0900	0.0500	1.15
f.	/ə/ - /ə/	0.6000	0.1500	0.1500	4.74
g.	/ə/ - /./	0.0200	0.1500	0.0500	1.42

In order to calculate the LD for /pe/ and /pelə/, the maximum PMI in the alignments is determined. In example (5) the maximum PMI equals 21.92. Subsequently the relevant pairings are subtracted from the maximal PMI and the resulting values are summed. For /pe/, this results in subtracting the values in rows a, b, c, e and g in example (5) from the maximum PMI and adding up the resulting values. For /pelə/, this results in adding lines a, b, c, d and f in example (5) after subtracting them from the maximal PMI. This results in a LD of 82.53 for /pe/ and 75.15 for /pelə/. This example shows that a word that is has a more identical alignment (/pelə/) has a smaller LD. Moreover, intuitively, /pelə/ is closer to the target than /pe/.

After the first pass through the transcript, the procedure is repeated. The Levenshtein algorithm is used to generate a new alignment, but this time with an alternative weighting scheme, viz. the dynamically computed cost model. In the first pass, matching segments received a “cost” of 0 and non-matching segments (insertions, substitutions or deletions) a “cost” of 1. From the second pass onwards, the segment distances computed during the previous pass are used in constructing a new alignment. This iterative process of aligning targets and renditions and computing segment distances, is stopped once two consecutive alignments are identical, and, hence, convergence is reached.

When the final cost model is computed, the LD between an adult target and a child’s rendition is the sum of distances between the individual segments. But since LD is relative to a word’s length, the LD score was normalized for word length by computing the average cost score per word (Heeringa, 2004). Applying this to the examples in (4) and (5), this means dividing the resulting LD of both child renditions by 5, i.e. the word length in phonemes. This eventually results in 16.51 for /pe/ and 15.03 for /pelə/. The rationale for this normalization is a follows: The score of a perfect match is most likely a non-zero distance. Consequently a short incorrectly produced word possibly receives a smaller LD value than a long perfectly produced word solely because the latter is longer. This undesirable effect of word length is prevented as far as possible by calculating the average LD.

Applying LD to transcriptions of children’s speech requires two more adaptations of the procedure for computing LD with a dynamic cost model. The first adaptation concerns the very first step in constructing the cost model dynamically. In constructing the cost model in the first pass, the cost model takes 0 for a segmental match and 1 for a mismatch, and subsequently PMI is computed. But suppose a child systematically deletes a particular segment in the renditions of adult targets. In that case the pairing of that segment with an empty segment will be highly frequent, and hence will be preferred to even the correct pairing. In order to circumvent this undesirable situation, the procedure was extended to a two-phase procedure. In the initial phase, a cost model of the adult language was constructed. For this purpose the transcription of the actual adult speech was aligned with a target transcription derived from the Fonilex database. The resulting cost model was computed for the entire corpus of adult speech and was considered to be a good estimate of adults' speech patterns. Subsequently, in the second phase, the children's speech was aligned with the adult targets. In the first pass the cost model derived from the adult transcriptions was used, so that situations such as a deletion being less costly than the adult model were avoided. But an implication of this restructuring of the procedure was that particular pairs of segments encountered in lining up child and adult transcriptions were not represented in the initial (adult) cost model. For instance, substitution of a closed front vowel by an open back vowel is hardly expected in adult language. Hence the pairing of /i/ with /a/ is not expected in the adult cost model, so its probability has to be estimated in some way. For this purpose the second adaptation of the procedure consisted of introducing Katz Smoothing (Chen & Goodman, 1998; Katz, 1987) before calculating the PMI. Through Katz Smoothing unobserved pairs of segments are assigned an estimated probability. They receive a small probability and thus a large cost (or distance) in the dynamic cost model.

Summing up, given the corpus of dyadic interactions, the adults' speech of the entire corpus was first used in order to construct a cost model that reflects the characteristics of adult spoken Dutch. Next, the Levenshtein distance was computed for each transcription of each individual child at the various ages under consideration.

Data analyses

Dynamic Levenshtein distance (LD) of each phonemic transcription of a word token was computed automatically after excluding substandard words for which there was no standard pronunciation in the Fonilex database. Phonemic accuracy development in terms of LD was investigated longitudinally between ages 1;2 and 2;0 for the NH children and between word onset after implantation up to age 2;0 for the children with CI**.** Phonemic accuracy was investigated cross-sectionally at ages 3;0, 4;0 and 5;0. Age ranges were 2;10 – 3;4 for analyses at age 3;0, 3;9 – 4;3 for analyses at age 4;0 and 4;11 – 5;3 for analyses at age 5;0. Outliers were determined by the interquartile rule and omitted from further analyses. All statistical analyses were done in JMP® Pro 11 by means of multilevel modeling (MLM). Multilevel models, also called hierarchical linear models, were used for the longitudinal analyses and fixed occasion multilevel models for the cross-sectional comparisons. The data of the present study are structured hierarchically into three levels: Individual words, various observations at consecutive ages and different children. These levels are nested. Words are nested within the different consecutive observations, which are nested within individual children. In other words, there is some variation in the present dataset resulting from the nesting of variables at different levels: At various ages, different words may be used and those may differ between children. In contrast to for instance ANOVA's, MLM takes this variation into account (Baayen, 2008; Woltman, Feldstain, MacKay, & Rocchi, 2012). Thus, MLM captures for instance the variation in the amount of word tokens and the variation in the sample sizes (i.e. different number of children in each group).

For the first research goal the phonemic accuracy of NH children and children with CI was examined longitudinally up age 2;0. The fixed effects were the children's hearing status (henceforth, HearingStatus), their ages in months (henceforth, Age), the length of the adult target words in syllables (henceforth, UtteranceSyllableLength), and the ratio of utterance length in phonemes of the target words over the utterance length in syllables of the target words (henceforth, WordComplexity). This last variable comprises the complexity of the target words: The ratio is higher in more complex words. Random intercepts and slopes were introduced to model the variation between children.

For the second research goal, cross-sectional comparisons between NH children and children with CI at ages 3;0, 4;0 and 5;0 were made. Fixed effects were HearingStatus, UtteranceSyllableLength and WordComplexity. At each age a random effect of child was included.

For the third research goal, the impact of age of implant activation and length of implant use was examined. Therefore, all available data of children with CI were used, i.e. also the monthly speech samples up to 2;6 years after implantation. Fixed effects were UtteranceSyllableLength, WordComplexity, length of implant use in months (henceforth, HearingAge) and age of CI activation in months (henceforth, CIactivation) as well as their interaction. Random intercepts and slopes were introduced to model inter-subject variation. In all analyses a significance level of $p<0.05$ was set.

Results

Phonemic accuracy development up to age 2;0

Analyses in this section discuss the longitudinal development of LD from word onset up to age 2;0 for NH children and children with CI. Table 4 displays the results of fitting the mixed effect model. The predicted values of the model are plotted in Figure 1. As Figure 1 demonstrates, LD of children with CI is higher than LD of NH children. This means that the accuracy of children with CI is lower than the accuracy of NH age-mates. However, no significant main effect of HearingStatus is found, meaning that although children with CI's accuracy is lower than that of NH children, and hence their LD is higher, this difference does not reach statistical significance. In addition, Figure 1 shows a linear decrease of LD with age, and as can be inferred from Table 4 the effect of Age is significant ($p = 0.0007$). No interaction effect between HearingStatus and Age is found, indicating that the development depicted in Figure 1 is highly similar for both groups of children.

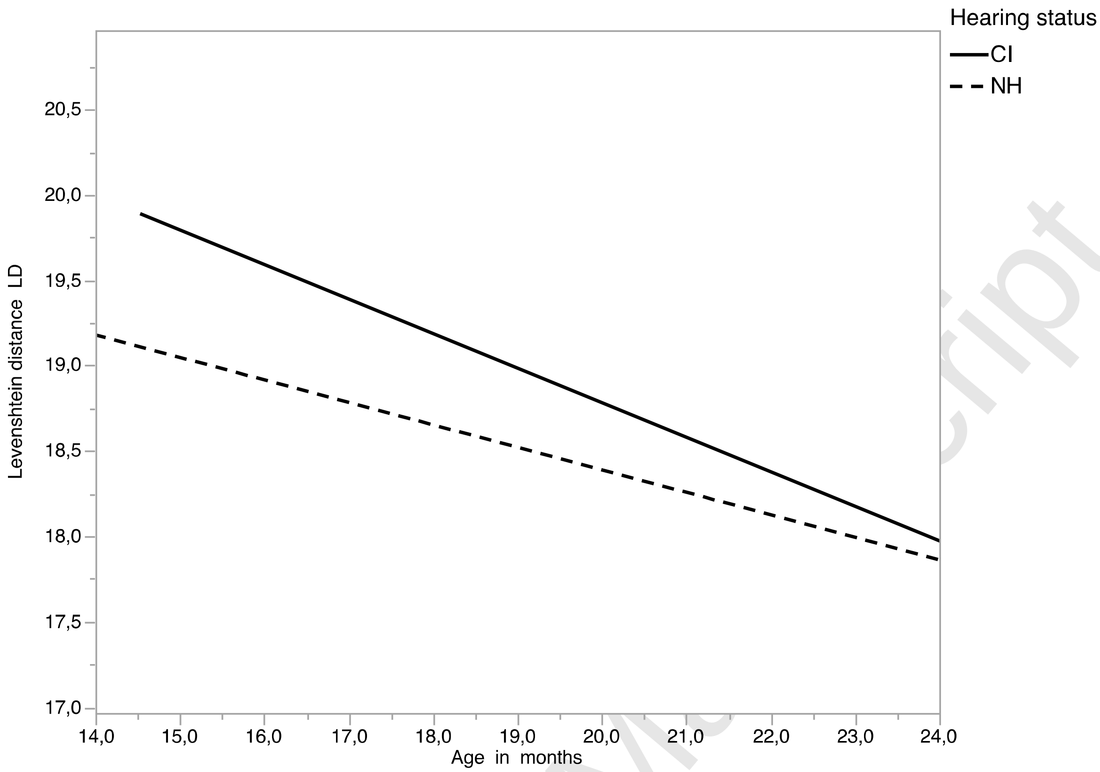

***Figure 1*. Development of Levenshtein distance of NH children and children with CI (Predicted values).**

Table 4 shows main effect results and the parameter estimates of UtteranceSyllableLength and WordComplexity. First of all, UtteranceSyllableLength is found to influence LD significantly. LD is higher with increasing UtteranceSyllableLength, thus accuracy is lower in longer words. Secondly, WordComplexity is found to influence LD as well: LD increases with increasing WordComplexity, which means that accuracy is lower in more complex words.

Importantly, some interaction effects of UtteranceSyllableLength and WordComplexity with Age and HearingStatus are found. First of all, Table 4 shows an interaction between both Age and UtteranceSyllableLength and Age and WordComplexity. This means that the influence of UtteranceSyllableLength and WordComplexity on LD decreases with time, or, in other words, that the increase of LD with increasing syllable length and word complexity becomes less steep as children grow older. Secondly, Table 4 shows that whereas there is no interaction between UtteranceSyllableLength and HearingStatus, there is an interaction between HearingStatus and WordComplexity. The predicted values of the model are plotted in Figure 2. It appears that the influence of UtteranceSyllableLength is the same for both groups of children, but WordComplexity influences LD differently in both groups of children. WordComplexity is found to influence LD more severely in children with CI than in NH children: The increase of LD with increasing WordComplexity is steeper in children with CI than in NH children.

***Table 4*. Fixed effect results and parameter estimates of the fixed effects up to age 2;0**

	df	Estimate	SE	F ratio	p
Intercept	1	18.0992	0.6072	/	<0.0001
Age	1	-0.0992	0.0283	12.27	0.0007
HearingStatus	1	0.9297	0.4860	3.66	0.0636
HearingStatus x Age	1	-0.0429	0.0226	3.59	0.0653
UtteranceSyllableLength	1	2.1255	0.1764	145.22	<0.0001
UtteranceSyllableLength x HearingStatus	1	-0.0462	0.0345	1.79	0.1806
UtteranceSyllableLength x Age	1	-0.0524	0.0080	42.94	<0.0001
WordComplexity	1	1.2746	0.1447	77.57	<0.0001
WordComplexity x HearingStatus	1	0.0912	0.0269	11.52	0.0007
WordComplexity x Age	**1**	-0.0232	0.0066	12.24	0.0005

df = degrees of freedom

SE = Standard error

NH is the reference category

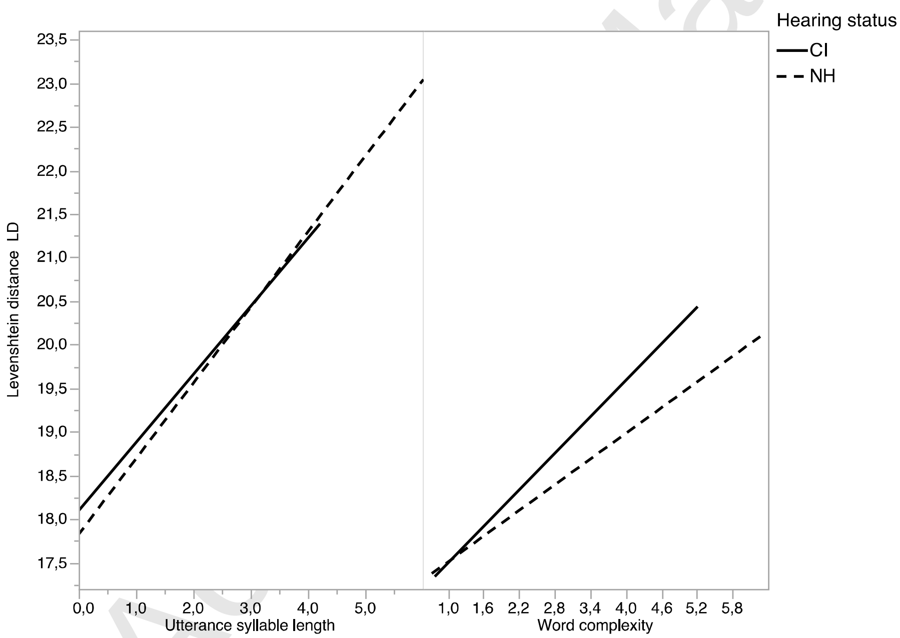

***Figure 2.* Interaction effect of HearingStatus with UtteranceSyllableLength (left pane) and WordComplexity (right pane) between ages 1;2 and 2;0 (Predicted values)**

Phonemic accuracy development from ages 3;0 to 5;0

The analyses in this section pertain to the longitudinal samples of the children with CI and to the cross-sectional samples of the NH children. Therefore, the results will be discussed for each age separately.

At age 3;0, all main effects except HearingStatus were significant, as indicated in Table 5. UtteranceSyllableLength as well as WordComplexity influence LD significantly: LD increases with increasing UtteranceSyllableLength and with increasing WordComplexity. Furthermore, the estimated LD is higher in children with CI as compared to NH children, but this effect is not statistically significant. Note however that significant interactions of HearingStatus with UtteranceSyllableLength and WordComplexity are found. Consequently, the increase of LD with increasing UtteranceSyllableLength and increasing WordComplexity is steeper in children with CI as compared to NH children. These effects are plotted in Figure 3.

Table 5. **Fixed effect results and parameter estimates of the fixed effects at age 3;0**

	df	**Estimate**	**SE**	**F ratio**	**p**
Intercept 1	1	15.7047	0.1535	/	<0.0001
HearingStatus	1	0.0898	0.1535	0.34	0.5620
UtteranceSyllableLength	1	0.4069	0.0358	129.27	<0.0001
UtteranceSyllableLength x HearingStatus	1	0.2058	0.0358	33.06	<0.0001
WordComplexity	1	0.3157	0.0330	91.76	<0.0001
WordComplexity x HearingStatus	1	0.1043	0.0330	10.01	0.0016

df = degrees of freedom

SE = Standard error

NH is the reference category

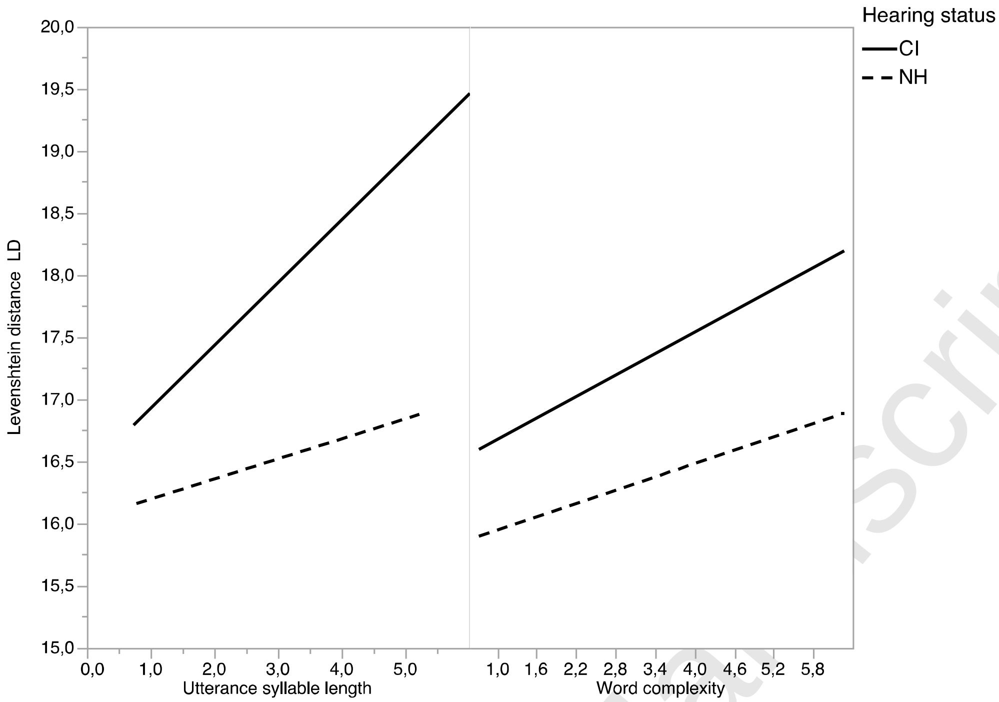

Figure 3. **Interaction effects of HearingStatus with UtteranceSyllableLength and WordComplexity on LD (age 3;0, predicted values)**

At age 4;0, the same trends emerge as at age 3;0. Table 6 presents the fixed effect results and the estimates of the fixed effect results. All effects are similar. No significant effect of HearingStatus is found. Furthermore, LD increases with increasing UtteranceSyllableLength and increasing WordComplexity. as was observed at age 3;0, the increase of LD with increasing UtteranceSyllableLength and increasing WordComplexity is higher in children with CI than in NH children. In Figure 4 the predicted LD values of the model are plotted.

Table 6. **Fixed effect results and parameter estimates of the fixed effects at age 4;0**

	df	Estimate	SE	F ratio	p
Intercept	1	15.4432	0.0537	/	<0.0001
HearingStatus CI	1	0.0427	0. 0537	0.63	0.4281
UtteranceSyllableLength	1	0.1798	0.0160	125.68	<0.0001
UtteranceSyllableLength x HearingStatus CI	1	0.0754	0. 0160	22.11	<0.0001
WordComplexity	1	0.1643	0.0143	131.44	<0.0001
WordComplexity x HearingStatus CI	1	0.0502	0. 0143	12.29	0.0005

df = degrees of freedom

SE = Standard error

NH is the reference category

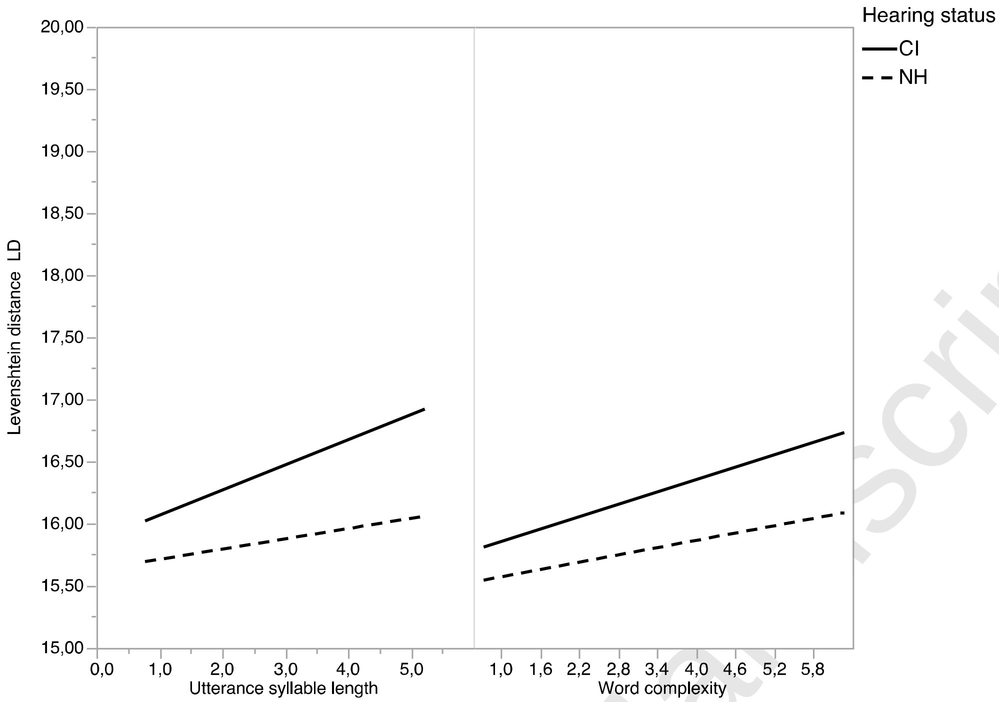

Figure 4. **Interaction effects of HearingStatus with UtteranceSyllableLength and WordComplexity on LD (age 4;0, predicted values)**

Fixed effect results and the estimates of the fixed effect results at age 5;0 are presented in Table 7. No effect of HearingStatus is found, but UtteranceSyllableLength as well as WordComplexity influence LD significantly. These effects are shown in Figure 5. In addition, no significant interaction between HearingStatus and UtteranceSyllableLength and between HearingStatus and WordComplexity are found. This means that WordComplexity and UtteranceSyllableLength influence LD in a comparative way in both groups of children.

Table 7. **Fixed effect results and parameter estimates of the fixed effects at age 5;0**

	df	Estimate	SE	F ratio	p
Intercept	1	15.3585	0.0337	/	<0.0001
HearingStatus	1	-0.0381	0.0337	1.27	0.2612
UtteranceSyllableLength	1	0.0682	0.0110	38.70	<0.0001
UtteranceSyllableLength x HearingStatus	1	0.0122	0.0110	1.23	0.2675
WordComplexity	1	0.0381	0.0097	15.39	<0.0001
WordComplexity x HearingStatus	1	-0.0053	0.0097	0.30	0.5864

df = degrees of freedom

SE = Standard error

NH is the reference category

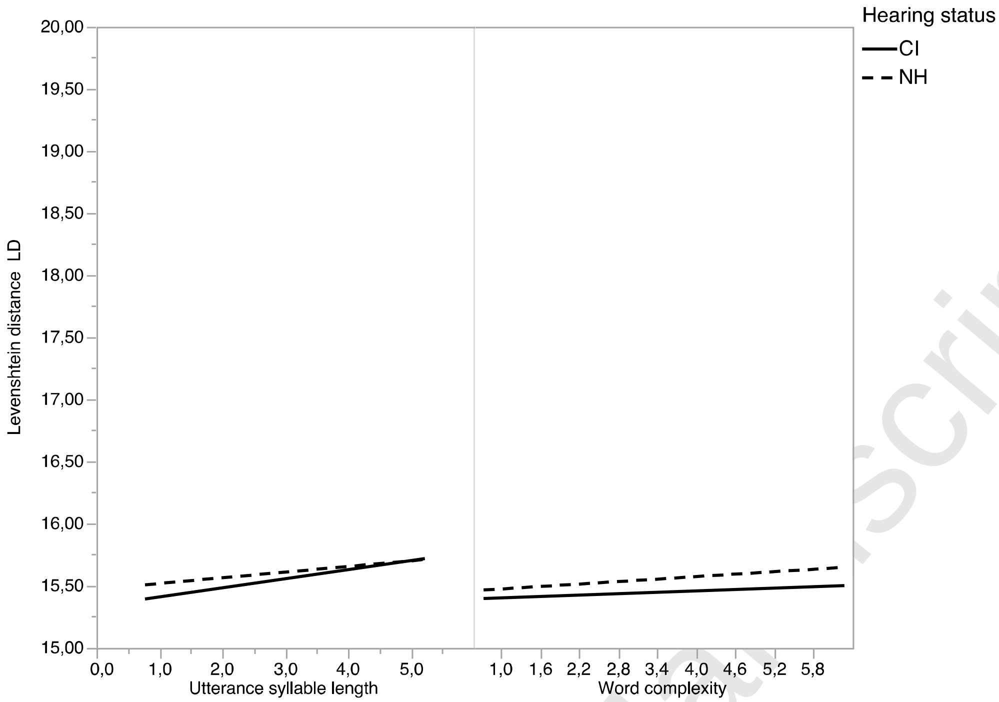

Figure 5. **Interaction effects of HearingStatus with UtteranceSyllableLength and WordComplexity on LD (age 5;0, predicted values)**

Influence of length of implant use and age at CI activation

In order to study the influence of age at implant activation and length of implant use the (longitudinal) data of the children with CI were analyzed separately, with HearingAge and age at CIactivation as predictors in the model (Table 8). LD is plotted relative to HearingAge in Figure 6. There is a quadratic, but overall decreasing effect of HearingAge on LD. Next, LD is higher in later implanted children, but this difference is not statistically significant, indicating that the effect of age at CIactivation on LD can safely be attributed to sampling error. No interaction effect between HearingAge and CIactivation is found, indicating that the decrease of LD in children with CI implanted at a later age is not significantly different from the decrease of LD in earlier implanted children.

UtteranceSyllableLength and WordComplexity significantly affect LD: LD increases with increasing UtteranceSyllableLength and increasing WordComplexity, as presented in Figure 7. For both variables, interactions with CI activation and HearingAge are found. The increase of LD with increasing WordComplexity is less steep in children with later CIactivation. The interaction of WordComplexity and CIactivation is significant, while the interaction of UtteranceSyllableLength and CIactivation is not. Next, the impact of both UtteranceSyllableLength and WordComplexity on LD decreases significantly with HearingAge.

***Table 8.* Fixed effect results and parameter estimates of the fixed effects for children with CI**

	df	Estimate	SE	F ratio	p
Intercept	1	16.4720	0.7527	/	<0.0001
HearingAge	1	-0.0535	0.0131	16.69	0.0039
HearingAge ^{2}	1	0.0010	0.0001	1267.55	<0.0001
CIactivation	1	0.0154	0.0531	0.08	0.7797
CIactivation x HearingAge	1	-0.0007	0.0009	0.68	0.4363
UtteranceSyllableLength	1	0.9946	0.0402	611.17	<0.0001
UtteranceSyllableLength x HearingAge	1	-0.0171	0.0007	640.03	<0.0001
UtteranceSyllableLength x CIactivation	1	-0.0031	0.0024	1.75	0.1860
WordComplexity	1	0.9502	0.0359	699.47	<0.0001
WordComplexity x HearingAge	1	-0.0203	0.0006	1339.48	<0.0001
WordComplexity x CIactivation	1	-0.0043	0.0020	4.42	0.0354

df = degrees of freedom

SE = Standard error

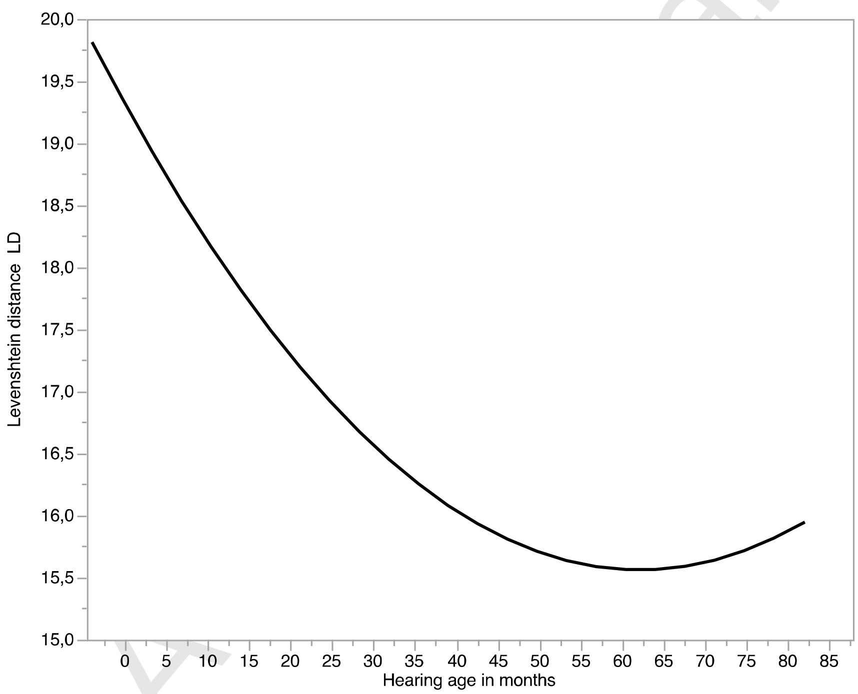

***Figure 6*. LD development of children with CI with HearingAge up to age 5;0 (Predicted values)**

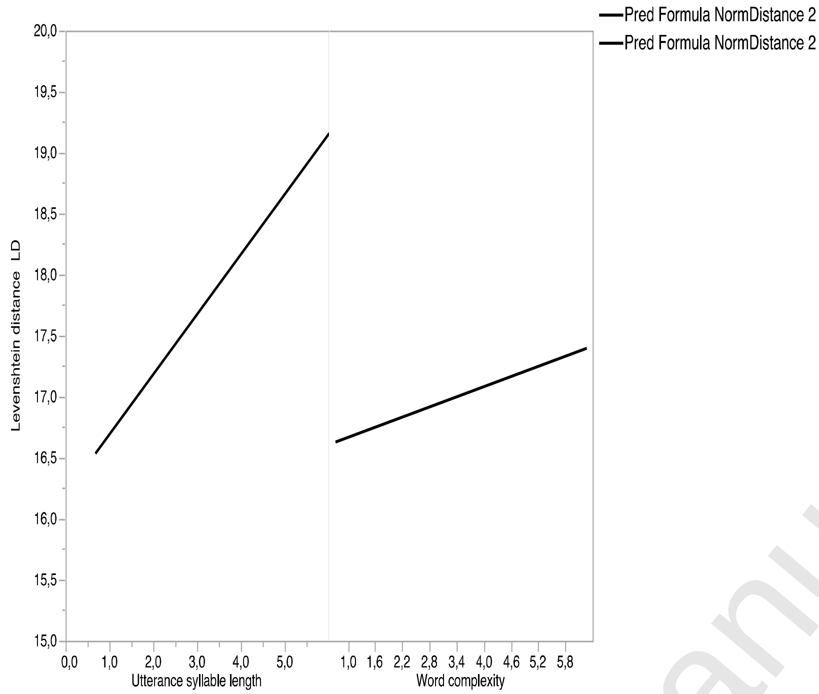

Figure 7. **Effects of UtteranceSyllableLength and WordComplexity on LD (children with CI) (Predicted values)**

Discussion

The aim of this study was to examine phonemic accuracy of Dutch-speaking children with CI in comparison to NH age-mates in order to determine if phonemic accuracy of children with CI is age appropriate. Accuracy development of early implanted children with CI was examined immediately after cochlear implantation up to age 2;0 and also the longer term evolution up to age 5;0 were scrutinized. Phonemic accuracy was operationalized as the dynamic Levenshtein distance (LD) between the child's spontaneous word productions and the attempted target words. In addition, for the children with CI the effect of length of implant use (or hearing age) and age of CI activation on LD were studied.

Development of accuracy in NH children and children with CI

How does phonemic accuracy develop in children with CI in comparison to NH children? The main result of our study, based on naturalistic longitudinal data, is that at the onset of word production, children with CI's renditions of adult words are consistently less accurate than those of NH age mates, but this difference is statistically not significant. This finding contrasts with reports in the literature: The phonemic accuracy of young children with CI has been shown to be significantly lower than the accuracy of NH children (Schauwers, Taelman, et al., 2008; Van den Berg, 2012; Warner-Czyz, 2005). Note that Warner-Czyz (2005) compared NH children and children with CI on their lexical age, i.e. up to six months after the onset of meaningful speech, while we compared NH children and children with CI on chronological age. A second main finding of the present study relates to the development of phonemic accuracy. At later ages and with prolonged device use production accuracy has been found to be lower in children with CI as compared to NH children (Eriks-Brophy et al., 2013; Ertmer & Goffman, 2011; Ertmer et al., 2012). Unlike these findings, no main effect of hearing status was found in the cross-sectional data analyzed in the present study at ages 3;0, 4;0 and 5;0, indicating that there is no statistically significant difference in phonemic accuracy of children with CI and age-matched NH children. Thus the outcomes of the present study agree with previous reports in the sense that phonemic accuracy is found to be lower in children with CI in comparison to NH age mates. But our findings are conflicting with the current literature in the sense that the difference in accuracy is not found to be statistically significant in the present study.

How can this discrepancy be explained? Obviously there are quite a few factors that influence the outcome of speech and language development after pediatric cochlear implantation, including child related factors (e.g., gender, the etiology of deafness, additional disabilities), audiological factors (e.g., bilateral auditory stimulation, either with a second CI or contralateral hearing aid) and environmental factors (e.g., communication mode, parental involvement in the rehabilitation), in addition to the variability among children that is intrinsic in the speech and language development process (Boons et al., 2013).

The difference between our results up to age 5;0 and the results of Eriks-Brophy et al. (2013), Ertmer and Goffman (2011) and Ertmer et al. (2012) may be due to a difference in mean age at implantation of the children with CI. Whereas the mean age at implantation was 2;0 (SD = 1;1) in Eriks-Brophy et al. (2013) and 1;6 (SD = 0;6) in Ertmer and Goffman (2011) and Ertmer et al. (2012), the mean age at implantation is considerably lower in the present article (1;0, SD = 0;5). Furthermore, Eriks-Brophy et al. (2013) compared NH children to a group of both children with CI and children with hearing aids. These methodological differences possibly explain why Eriks-Brophy et al. (2013), Ertmer and Goffman (2011) and Ertmer et al. (2012) found a main effect of hearing status up to age 5;0 and we did not. However, as the same participants with CI were studied in Schauwers, Taelman, et al. (2008), Van den Berg (2012) and the present paper, age at implantation cannot explain the discrepant findings.

There are however several other important methodological differences between the present study and the studies reported in the literature. First of all, the speech samples analyzed in order to establish phonemic accuracy vary from rigidly elicited speech to completely unrestrained spontaneous speech. In the present study, spontaneous speech is used. Eriks-Brophy et al. (2013), Ertmer and Goffman (2011) and Ertmer et al. (2012) studied elicited speech using standardized tests. It is unclear what the exact impact of the speech sampling context is for phonemic accuracy in young children, be it NH children or children with CI. Our study reveals that in spontaneous speech samples, the children with CI are less accurate than their NH age-mates, however the difference is statistically not significant and hence may be due to "sampling error". However the studies of Ertmer and colleagues seem to indicate that in elicited speech (in a clinical context) children with CI are outperformed by NH children. This apparent opposition is in need of further clarification.

The most important methodological difference between our results and the results of Eriks-Brophy et al. (2013), Ertmer and Goffman (2011), Ertmer et al. (2012), Schauwers, Taelman, et al. (2008), Van den Berg (2012) and Warner-Czyz (2005) is the applied measure of accuracy. In most cases a standardized accuracy measure was used. Eriks-Brophy et al. (2013) used the GFTA-2. Ertmer and Goffman (2011) as well as Ertmer et al. (2012) calculated PCC, and Schauwers, Taelman, et al. (2008) applied pMLU. As indicated in the introduction PCC and pMLU are no optimal measures of accuracy for several reasons. Warner-Czyz (2005) evaluated consonants based on their agreement in manner and place, but not in voice and evaluated accuracy of vowels based on their agreement in height and backness. Van den Berg (2012) used a restrictive accuracy measure: A word was considered to be incorrect even if only one phoneme deviated from the target. Hence her measure was overly conservative in computing phonemic accuracy. In contrast, in the present study Levenshtein distance is used, a much more fine-grained measure: Unlike pMLU and PCC, LD penalizes all types of speech errors, including insertions. In addition, LD considers the frequency of speech errors or deviations. Therefore, frequent and less severe speech errors are penalized less than infrequent and severe speech errors. The use of this more refined measure of accuracy arguably influences the accuracy results of children with CI and NH children.

To recapitulate, even though the same participants with CI were studied in Schauwers, Taelman, et al. (2008), Van den Berg (2012) and the present study, different results in accuracy up to age 2;0 are found. In addition, our results up to age 5;0 also differ from Eriks-Brophy et al. (2013), Ertmer and Goffman (2011) and Ertmer et al. (2012). In all mentioned studies, differences between NH children and children with CI are found, while in our study, no main effect of hearing status is found. This discrepancy can be explained by methodological differences, including the language studied, the type of speech studied, age at implantation of children with CI, statistical methods used to map variation, the number of NH children in the control groups and the period studied and the applied measure of accuracy.

Our results have to be considered with some care, as relatively limited speech samples were available for a relatively limited number of children with CI. Even though the use of multilevel modeling takes into account variation in the data, such as the unequal number of children in each group and the unequal amount of word tokens. Obviously, more word tokens and more children with CI would increase the accuracy of the estimates. In addition, a phonemic broad transcription is applied in the present paper. Our results indicate that at the phonemic level children with CI become as accurate as children with normal hearing. But a broad phonemic transcription does not take into account fine phonetic variation and articulatory distortions, such as lisps. Adding such phonetic detail to the transcriptions to the computation of the LD measure would add a phonetic layer to the present study, which was not the current aim. A narrow phonetic transcription could fine-tune the LD calculations.

But, a narrow transcription is evidently much more time consuming than a broad transcription. This implies that in the same amount of transcription time, a lower number of productions can be transcribed, which in turn would reduce the precision of the estimates in the statistical analyses. Hitting the balance between time investment and amount of material to be analyzed is a delicate matter in this respect. In the present study broad phonemic transcriptions were used, leaving the unexplored phonetic variation on our research agenda.

Effect of target word length and target word complexity

Our results suggest effects of both target word complexity and target word syllable length in the NH children and children with CI. Target word complexity and target word syllable length influence accuracy significantly: Phonemic accuracy decreases with increasing target word syllable length and increasing target word complexity. Thus, the more syllables in the target word, the less accurate the child's rendition, and, the more complex the target word, for instance caused by the presence of consonant clusters, the less accurate the child's rendition of that word. This is in line with the literature on children's repetitions of nonwords, both in NH children and in children with CI (Burkholder-Juhasz et al., 2007; Gathercole et al., 1991; Macrae, 2013; Nittrouer et al., 2014; Von Mentzer et al., 2015), showing that accuracy decreases with increasing syllable length and with increasing phonological complexity of the target nonwords. In contrast to these studies, our results are not based on nonword repetition but on spontaneous speech. Nevertheless, our results are highly similar. The length of the target word is decisive, as phonemic accuracy decreases with increasing number of syllables in the target word. The same is true for target word complexity: Accuracy decreases when word complexity increases. Our results show that the impact of target word complexity and target word syllable length decreases in NH children. A longitudinal analysis of children with CI showed comparable results. The impact of target word complexity and target word syllable length decreases when children with CI have more hearing experience. Thus, initially, target word complexity and target word syllable length have a significant, negative influence on accuracy. Even though this influence is still significant at age 5;0, our results suggest that the influence diminishes.

Interestingly, the syllable length of target words and their complexity influence phonemic accuracy more dramatically in children with CI as compared to NH children. Consequently, the phonemic accuracy of children with CI is lower than the phonemic accuracy of NH children when producing longer and more complex words. Thus children with CI are found to be less accurate than NH children. However, our results suggest that this difference is subtler than a main effect of hearing status as reported in Eriks-Brophy et al. (2013), Ertmer et al. (2012), Schauwers, Taelman, et al. (2008) and Van den Berg (2012). Our analyses indicate that not hearing status as such, but the interaction between hearing status and target word complexity and target word length is fundamental to understand the difference between NH children and children with CI. The difference between the two groups is to be found in the production of longer and more complex words and not so much in short and simple words.

Between ages 1;2 and 2;0, the decrease of phonemic accuracy with increasing word complexity is more outspoken in children with CI as compared to NH children. This complements the findings of Schauwers, Taelman, et al. (2008) that children with CI target less complex words than NH children: The pMLU of the target words is significantly lower in children with CI. In a similar vein, Van den Berg (2012) showed that children with CI use relatively more monosyllabic words than NH children in their second year of life. Thus they acquire relatively more simple and less complex words, and with increasing complexity their accuracy is significantly lower than the accuracy of NH children.

At ages 3;0 and 4;0, the interactions between hearing status and target word length and complexity remain. Phonemic accuracy decreases more dramatically in children with CI as compared to NH children with increasing target word complexity and increasing target word syllable length respectively. Phonological processing of more complex words is thus more difficult for children with CI up to age 4;0. However the difference between the two groups has disappeared by age 5;0. By that age children with CI appear to have caught up with their NH peers. This striking conclusion holds for phonemic accuracy and is highly similar to Nicholas and Geers (2007) who concluded that children with CI have caught up with their NH peers at age 4;6 for varying measures for vocabulary, sentence complexity and morphology. For Dutch children with CI, this is also true for phonemic accuracy.

Effect of length of implant use and age at implantation

Speech and language development of children with CI is often reported to depend on the age at implantation (Nicholas & Geers, 2007; Schauwers, Taelman, et al., 2008; Van den Berg, 2012) and length of device use (Blamey et al., 2001; Tomblin et al., 2008). Our results suggest that the age at implant activation has no significant effect on phonemic accuracy. This might be due to the fact that all participants in this study were implanted at a very young age, namely before age 1;8. This finding is in line with Szagun and Stumper (2012) who concluded that language development of children with CI implanted before the end of what they call “the sensitive period” is not significantly affected by the age of implantation. In contrast, length of implant use influences language development in early implanted children with CI (Szagun & Stumper, 2012). In agreement with Szagun and Stumper (2012), our results suggest that length of implant use influences the development of phonemic accuracy significantly. This is also in line with the focus on length of implant use in other long-term effect studies (Blamey et al., 2001; Tomblin et al., 2008), even though the participants in these studies were implanted after the age of 2;0. Thus, for young implanted children with CI, accuracy is susceptible to the length of implant use and not so much to the age of implant activation. But in the group of children with CI investigated in the present study (age at activation between ages 0;6 and 1;8), the influence of word complexity is smaller in children with CI with later implant activation.

Conclusion: clinical and theoretical implications

Our results suggest that for children with CI the accurate production of more complex and longer words are especially problematic in comparison with their NH age-mates up to age 4;0. Children with CI are still found to be delayed with respect to the phonemic accuracy of long and complex words. Their accuracy is comparable to NH peers but only for words of restricted length. This suggests a focus for speech and language therapy: with more hearing experience, children with CI appear to master shorter words well at a segmental level, but longer, more complex words remain problematic. Hence, clinical intervention might take up this finding and integrate more material varying in length and complexity, provided that is not the case yet. Rehabilitation programs should also focus on the underlying factors accounting for the observation that phonemic accuracy deteriorates with increasing word length and complexity. In the literature some possible causes have been suggested which may be an additional focus of rehabilitation programs. First of all, Houston and Bergeson (2014) showed that children with CI are less attentive to speech sounds in the ambient language. Moreover they have a delay in integrating audiovisual information (Houston et al. 2012). As a result, they are found to have poorer phonological awareness and less precise phonological representations (Lund, Werfel, & Schuele, 2015). Consequently, children with CI may benefit from an enhancement of their attention to speech and to speech related information, including audiovisual correlates, as this may result in better phonological awareness and representations. Secondly, children with CI are typically found to have poorer phonological short-term working memory, which results in poorer performances in more complex tasks, such as the repetition of an increasing number of digits (Kronenberger, Beer, Castellanos, Pisoni, & Miyamoto, 2014; Pisoni & Cleary, 2004; Pisoni, Kronenberger, Roman, & Geers, 2010). A similar effect is found in the present paper. Children with CI produce longer and more complex words less accurately. In other words, enhancing the phonological working memory skills of children with CI as part of their rehabilitation may be beneficial for their processing of longer and more complex words.

To conclude, the present study has shown that the accuracy of children with CI is lower as compared to NH peers, but only in interaction with target word complexity and target word syllable length. This reveals the importance of including explanatory variables when comparing both groups of children. In other words, it seems relevant to compare NH children and children with CI not only as such, but it seems also relevant to include interactions with other variables. Moreover, target word complexity and target word length are also shown to influence accuracy of NH children. This suggests that similar factors influence accuracy in NH children and children with CI, even though the effect is more outspoken in children with CI. Thus, target word complexity and syllable length of the target word are decisive factors in accurate producing words. Less complex and shorter words are produced more accurately than more complex and longer words. This is in line with other literature on factors influencing accuracy. As shown, those factors influence the accuracy of children with CI more severely as compared to NH peers. Furthermore, our results showed that accuracy is more susceptible to hearing age than to age of implant activation in early implanted children with CI. Thus, length of implant use can be seen as a good predictor variable of accuracy development in children with CI.

Acknowledgments

This work was supported by grant G.0138.13 of the FWO in Flanders. We would like to thank I. Molemans, K. Schauwers, K., R. van den Berg and L. Van Severen for collecting and transcribing the data. We would also like to thank S. De Maeyer for his advice on the statistical analyses. Our special thanks go to M. Pettinato for her critical comments on a previous version of this paper.

References

Altvater-Mackensen, N., & Fikkert, P. (2010). The acquisition of the stop-fricative contrast in perception and production. *Lingua,* 120, 1898 - 1909. doi: 10.1016/j.lingua.2010.02.010

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

Beers, M. (1995). *The phonology of normally developing and language-impaired children.* Amsterdam University, Unpublished doctoral dissertation.

Blamey, P., Barry, J., Bow, C., Sarant, J., Paatsch, L., & Wales, R. (2001). The development of speech production following cochlear implantation. *Clinical Linguistics & Phonetics,* 15(5), 363 - 382. doi: 10.1080/02699200010017823

Boons, T., De Raeve, L., Langereis, M., Peeraer, L., Wouters, J., & Van Wieringen, A. (2013). Expressive vocabulary, morphology, syntax and narrative skills in profoundly deaf children after early cochlear implantation. *Research in developmental disabilities*, 34, 2008 - 2022. doi: 10.1016/j.ridd.2013.03.003

Burkholder-Juhasz, R., Levi, S., Dillon, C., & Pisoni, D. (2007). Nonword repetition with spectrally reduced speech: some development and clinical findings from pediatric cochlear implantation. *Journal of Deaf Studies and Deaf Education,* 12(4), 472 - 485. doi: 10.1093/deafed/enm031

Carter, A., Dillon, C., & Pisoni, D. (2002). Imitation of nonwords by hearing impaired children with cochlear implants: suprasegmental analyses. *Clinical Linguistics & Phonetics*, 16(8), 619 - 638.

Chen, S., & Goodman, J. (1998). An empirical study of smoothing techniques for language modeling: Technical report, Harvard university.

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(6), 628 - 644. doi: 10.1097/01.aud.0000240640.59205.42

Cucchiarini, C. (1996). Assessing transcription agreement: methodological aspects. *Clinical Linguistics & Phonetics*, 10(2), 131 - 155.

Desloovere, C., Verhaert, N., Van Kerschaver, E., & Debruyne, F. (2013). Fifteen years of early hearing screening in Flanders: impact on language and education. *B-ENT,* Suppl 21, 81 - 90.

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: a normative study of British English-speaking children. *Clinical Linguistics & Phonetics,* 17(8), 617 - 643. doi: 10.1080/0269920031000111348

Elffers, C., Van Bael, C., & Strik, H. (2005). ADAPT: algorithm for dynamic alignment of phonetic transcriptions. Nijmegen: Departement of Language & Speech, Radboud University. Retrieved from http://lands.let.ru.nl/literature/elffers.2005.1.pdf

Eriks-Brophy, A., Gibson, S., & Tucker, S. (2013). Articulatory error patterns and phonological process use of preschool children with and without hearing loss. *The Volta Review*, 113(2), 87-125.

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. doi: 10.1044/1092-4388(2010/09-0165)

Ertmer, D., Kloiber, D., Jung, J., Kirleis, K., & Bradford, D. (2012). Consonant production accuracy in young cochlear implant recipients: developmental sound classes and word position effects. *American Journal of Speech-Language Pathology,* 21, 342 - 353. doi: 10.1044/1058-0360(2012/11-0118)

Gantz, B., Tyler, R., Woodworth, G., Tye-Murray, N., & Fryauf-Bertschy, H. (1994). Results of multichannel cochlear implants in congenital and acquired prelingual deafness in children: five-year follow-up. *The American Journal of Otology,* 15(s2), 1 - 7.

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(3), 349 - 367.

Gillis, S., & Schaerlaekens, A. (Eds.). (2000). *Kindertaalverwerving. Een handboek voor het Nederlands*. Groningen: Martinus Nijhoff.

Heeringa, W. (2004). *Measuring dialect pronunciation differences using Levenshtein Distance.* Groningen University.

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. doi: 10.1016/j.lingua.2013.08.001

Ingram, D. (2002). The measurement of whole-word productions. *Journal of Child Language,* 29, 713-733. doi: 10.1017/S0305000902005275

Irwin, J., & Wong, S. (1983). *Phonological development in children 18 to 72 months*. Illinois: Southern Illinois University Press.

Johnson, W., & Reimers, P. (2010). *Patterns in child phonology*. Edinburgh: Edinburgh University Press

Jongstra, W. (2003). *Variation in reduction strategies of Dutch word-initial consonant clusters.* Toronto University, Unpublished doctoral dissertation.

Jusczyk, P. (1992). Developing phonological categories from the speech signal. In C. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological Development: Models, Research and Implications*. Maryland: York Press Timonium.

Katz, S.M. (1987). Estimation of probabilities from sparse data for the language model component of a speech recognizer. *IEEE Transactions on Acoustics, Speech, and Signal Processing,* 35, 400 - 401.

Kronenberger, W., Beer, J., Castellanos, I., Pisoni, D., & Miyamoto, R. (2014). Neurocognitive risk in children with cochlear implants. *JAMA Otolaryngology - Head & Neck Surgery,* 140(7), 608 - 615. doi: 10.1001/jamaoto.2014.757

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics,* 39, 159 - 174.

Lund, E., Werfel, K., & Schuele, M. (2015). Phonological awareness and vocabulary performance of monolingual and bilingual preschool children with hearing loss. *Child Language Teaching and Therapy,* 31(1), 85 - 100. doi: 10.1177/0265659014531261

Macrae, T. (2013). Lexical and child-related factors in word variability and accuracy in infants. *Clinical Linguistics & Phonetics,* 6-7, 497 - 507. doi: 10.3109/02699206.2012.752867

MacWhinney, B. (2000). *The CHILDES Project: tools for analyzing talk*. Mahwah: NJ: Lawrence Erlbaum Associates.

Mertens, P. (2001). Fonilex.   Retrieved 14 august, 2014, from http://bach.arts.kuleuven.be/fonilex/

Molemans, I., Van den Berg, R., Van Severen, L., & Gillis, S. (2012). How to measure the onset of babbling reliably? *Journal of Child Language,* 39(3), 523 - 552. doi: 10.1017/S0305000911000171

Nerbonne, J., & Heeringa, W. (2010). Measuring dialect differences. In J. Schmidt & P. Auer (Eds.), *Language and space: theories and methods in series handbooks of linguistics and cCommunication science* (pp. 550 - 567). Berlijn: Mouton De Gruyter,.

Nicholas, J., & Geers, A. (2007). Will they catch up? The role of age at cochlear implantation in the spoken language development of children with severe to profound hearing loss. *Journal of Speech, Language and Hearing Research,* 50(4), 1048 - 1062. doi: 10.1044/1092-4388(2007/073)

Nittrouer, S., Caldwell-Tarr, A., Sansom, E., Twersky, J., & Lowenstein, J. (2014). Nonword repetition in children with cochlear implants: a potential clinical marker of poor language acquisition. *American Journal of Speech-Language Pathology,* 23, 679 - 695. doi: 10.1044/2014_AJSLP-14-0040

Pisoni, D., & Cleary, M. (2004). Learning, memory and cognitive processes in deaf children following cochlear implantation. In F. G. Zeng, A. N. Popper & R. R. Fay (Eds.), *Cochlear Implants: Auditory Prostheses and Electric Hearing*. New York: Springer.

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(1), 60S - 74S. doi: 10.1097/AUD.0b013e3181ffd58e

Riches, N.G., Loucas, T., Baird, G., Charman, T., & Simonoff, E. (2011). Non-word repetition in adolescents with specific language impairment and autism plus language impairments: a qualitative analysis. *Journal of Communication Disorders,* 44, 23 - 36. doi: 10.1016/j.jcomdis.2010.06.003

Saaristo-Helin, K. (2009). Measuring phonological development: a follow-up study of five children acquiring Finnish. *Language and Speech,* 52(1), 55 - 77.

Saaristo-Helin, K., Savinainen-Makkonen, T., & Kunnari, S. (2006). The phonological mean length of utterance: methodological challanges from a crosslinguistic perspective. *Journal of Child Language,* 33, 179 - 190. doi: 10.1017/S0305000905007294

Sanders, N., & Chin, S. (2009). Phonological distance measures. *Journal of Quantitative Linguistics,* 16(1), 69 - 114. doi: 10.1080/09296170802514138

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

Schauwers, K., Gillis, S., & Govaerts, P. (2008). The Characteristics of prelexical babbling after cochlear implantation between 5 and 20 months of age. *Ear and Hearing,* 29(4), 627 - 637. doi: 10.1097/AUD.0b013e318174f03c

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: Cambridge Scholars Publishing.

Shriberg, L., Austin, D., Lewis, D., McSweeny, J., & Wilson, D. (1997). The percentage of consonants correct (PCC) metric: extensions and reliability data. *Journal of Speech, Language, and Hearing Research,* 40(4), 708 - 722.

Shriberg, L., & Kwiatkowski, J. (1982). Phonological disorders III: a procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders,* 47(256 - 270).

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language,* 38, 1 - 34. doi: 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). Blackwell: Blackwell Publishing.

Swerts, M., Kloots, H., Gillis, S., & De Schutter, G. (2003). *Vowel reduction in spontaneous spoken Dutch.* Proceedings of the ISCA & IEEE Workshop on spontaneous speech processing and recognition (31 - 34), Tokyo, Japan.

Szagun, G., & Stumper, B. (2012). Age or experience? The influence of age at implantation and social and linguistic environment on language development in children with cochlear implants. *Journal of Speech, Language and Hearing Research,* 55, 1640 - 1654. doi: 10.1044/1092-4388(2012/11-0119)

Taelman, H., Durieux, G., & Gillis, S. (2005). Notes in Ingram's phonological measures. *Journal of Child Language,* 32, 391 - 405. doi: 10.1017/S0305000904006774

Tobey, E., Geers, A., Brenner, C., Altuna, D., & Gabbert, G. (2003). Factors associated with development of speech production skills in children implanted by age five. *Ear and Hearing,* 24(1S), 36S - 45S. doi: 10.1097/01.AUD.0000051688.48224.A6

Tomblin, B., Spencer, L., & Lu, N. (2008). Long-term trajectories of the development of speech sound production in pediatric cochlear implant recipients. *Journal of Speech, Language, and Hearing Research,* 51, 1353 - 1368. doi: 10.1044/1092-4388(2008/07-0083)

Tye-Murray, N., Spencer, L., & Woodworth, G. (1995). Acquisition of speech by children who have prolonged coclear implant experience. *Journal of Speech and Hearing Research,* 38(2), 327 - 337.

Tyler, A., & Lewis, K. (2005). Relationships among consistency/variability and other phonological measures over time. *Topics in Language Disorders,* 25(3), 243 - 253.

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.* University of Antwerp, Unpublished doctoral dissertation.

Van Severen, L., Gillis, J., Molemans, I., Van den Berg, R., De Maeyer, S., & Gillis, S. (2013). The relation between order of acquisition, segmental frequency and function: the case of word-initial consonants in Dutch. *Journal of Child Language,* 40, 703 - 740. doi: 10.1017/S0305000912000219

Vihman, M., & McCune, L. (1994). When is a word a word? *Journal of Child Language,* 21, 517 - 542.

Von Mentzer, C., Lyxell, B., Sahlén, B., Dahlström, O., Lindgren, M., Ors, M., Kallioinen, P., Engström, E., & 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(3), 216 - 235. doi: 10.3109/02699206.2014.987926

Wagner, R.A., & Fisher, J.M. (1974). The string-to-string correction problem. *Journal of the ACM,* 21(1), 168 - 173.

Warner-Czyz, A. (2005). *The development of accuracy in ealry speech acquisition: relative contributions of production and auditory perceptual factors.* Univeristy of Texas, Austin.

Wieling, M., Margaretha, E., & Nerbonne, J. (2011). Inducing phonetic distances from dialect variation. *Computational Linguistics in the Netherlands Journal,* 1, 109 - 118.

Wieling, M., Margaretha, E., & Nerbonne, J. (2012). Inducing a measure of phonetic similarity from pronunciation variation. *Journal of Phonetics,* 40(2), 307 - 314. doi: 10.1016/j.wocn.2011.12.004

Wieling, M., Prokic, J., & Nerbonne, J. (2009). *Evaluating the pairwise string alignment of pronunciations*. Paper presented at the Proceedings of the 12th Meeting of the European Chapter of the Association for Computational Linguistics, Athens, Greece.

Woltman, H., Feldstain, A., MacKay, C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in quantitative methods for psychology,* 8(1), 52 - 69.

Learning outcomes

This article informs the reader about phonemic accuracy development in children. The reader will be able to (a) discuss different metrics to measure phonemic accuracy development, (b) discuss phonemic accuracy of children with CI up to five years of age and compare them with NH children, (c) discuss the influence of target word's complexity and target word's syllable length on phonemic accuracy, (d) discuss the influence of hearing experience and age of implantation on phonemic accuracy of children with CI.

Highlights

Levenshtein distance is a new measure of accuracy in child language research.
Target word complexity and syllable length affect phonemic accuracy.
Complexity and syllable length influence cochlear implanted children more severely.
Cochlear implanted children catch up on their normally hearing peers by age five.
Length of implant use influences accuracy of cochlear implanted children.