The processing of Dutch prosody with cochlear implants and vocoder simulations

Velde, D.J. van de

Author: Velde, D.J. van de
Title: The processing of Dutch prosody with cochlear implants and vocoder simulations
Issue Date: 2017-07-05

Chapter **2**

Basic measures of prosody in spontaneous speech of children with early and late cochlear implantation

Abstract

Research on prosody in speech produced by children with cochlear implants (CI) has revealed deviations from the speech of normally hearing (NH) peers, such as a high fundamental frequency (F0), elevated jitter and shimmer, and inadequate intonation. However, three important dimensions of prosody (temporal, intensity, and spectral) have not been systematically investigated or compared in production research. Given that in general the resolution in CI hearing is best for the temporal, followed by the intensity, and worst for the spectral dimension, we may expect that this hierarchy is also present in the speech production.

9 Dutch Early Implanted (EI), 9 Late Implanted (LI; division at 2 years of age) children and 12 hearing age matched NH controls were tested at 18, 24, and 30 months after implantation (CI) or birth (NH).

We expected that (1) there would be differences between CI recipients and controls on prosodic speech measures, (2) they would be smallest for temporal measures, followed by intensity measures and largest for

spectral measures, (3) they would be larger for later than for earlier implanted children (4) and they would diminish with increasing device experience.

From spontaneous speech data, 1,937 utterances were extracted. Of these utterances, nine outcome measures along the spectral, intensity and temporal dimensions were subjected to Principle Component Analysis (PCA) and, using Linear Mixed Modelling, compared between Group, Session, and Gender, as well as their interactions.

PCA combined three measures into one, leaving three temporal and three spectral measures. On most measures, interactions of Group and/or Gender with Session were significant. For CI recipients as compared to controls, performance on temporal measures was not in general more deviant than spectral measures, although differences were found for individual measures. LI had a tendency to be closer to NH than EI. Groups converged over time.

The hypothesis regarding differential deviations for the different phonetic dimensions was not supported. This suggests that the appropriateness of the production of basic prosodic measures does not depend on auditory resolution. Rather, it seems to depend on the amount of control necessary for speech production. Chronological age, hearing status and gender of the speaker influence the development of the measures.

2.1 Introduction

Most people who suffer from severe or profound hearing loss are nowadays treated with cochlear implantation (CI), which partly restores their hearing. Despite major advantages in spoken communication relative to pre-implantation, the CI recipients’ hearing situation is not like that of normally-hearing (NH) people. Characteristics of the device and the CI recipient’s auditory history limit, in particular, the perception of speech prosody (Meister et al., 2007), music (Looi, Gfeller, & Driscoll, 2012) and hearing in noise (Friesen, Shannon, Baskent, & Wang, 2001). This hearing situation does not only affect perception of speech, but is expected to result in deviant speech output as well, since there is a link between hearing capacity and speech production performance, i.e., self-monitoring of speech (Guenther, 2006; Levelt, 1983).

The speech of CI recipients has been investigated by at least two different types of studies. The first type (which can be called the ‘normative’ type) is to compare CI recipients’ voices at one or more moments in time after implantation to their pre-implantation voices and/or to the voices of normally hearing peers, as part of the same study or as normative data from previous research (Evans & Deliyski, 2007; Goffman, Ertmer, & Erdle, 2002; Lane et al., 1998; Perrin, Berger-Vachon, Topouzkhanian, Truy, & Morgon, 1999; Seifert et al., 2002; Ubrig et al., 2011; Uchanski & Geers, 2003; Valero Garcia, Rovira, & Sanvicens, 2010). The second type of research (the ‘on/off’ type) involves a comparison between the performance of (more or less experienced) CI users in a condition in which their implant is temporarily turned off and one in which it is turned on again (Higgins, McCleary, & Schulte, 2001; Poissant, Peters, & Robb, 2006; Tye-Murray, Spencer, Bedia, & Woodworth, 1996).

Outcomes across studies of both types vary considerably, both in the direction and the amount of deviations (if any) from the norm. This variability has been attributed to the divergence in the following methodological factors: speech material (sustained vowels, syllables, read-aloud continuous speech or spontaneous speech), assessment techniques (aerodynamic/physiologic, standard acoustic analysis, custom-made acoustic analysis or perceptual evaluation), age of the participants, speech-processing strategy of the implant and age of implant activation (Baudonck, van Lierde, Dhooge, & Corthals, 2011). The lack of convergence in the results so far is substantiated by a review of 27 articles about the voice quality of CI users (Coelho, Brasolotto, & Bevilacqua, 2012), which concluded that the number of effective studies is too small to draw clear conclusions.

Nevertheless, a number of impressionistic generalizations about voice and speech measures can be made from the pooled investigations on CI users with varying hearing histories so far. The fundamental frequency (F0) is high before implantation, on normative type studies (Oster, 1987; Perkell, Lane, Svirsky, & Webster, 1992; Szyfter et al., 1996; Ubrig et al., 2011) or when the implant is turned off, i.e., in on/off type studies (Monini, Banci, Barbara, Argiro, & Filipo, 1997; Poissant et al., 2006; Svirsky, Lane, Perkell, & Wozniak, 1992), and drops gradually after implantation. Variability of F0, or vF0 (Ball & Ison, 1984; Holler et al., 2010; Ubrig et al., 2011), and jitter (Fourcin, Abberton, Richardson, & Shaw, 2011; Hocevar-Boltezar et al., 2006) decrease after implantation. The nasal resonance of the speech is in general either too low (Monini et al., 1997; van Lierde, Vinck, Baudonck, De Vel, & Dhooge, 2005) or too high (Hassan et al., 2011a; Nguyen, Allegro, Low, Papsin, & Campisi, 2008; Svirsky, Jones, Osberger, & Miyamoto, 1998; Ubrig et al., 2011), but interacts with the principal resonance cavity of the sound (Baudonck, van Lierde, D'Haeseleer, & Dhooge, 2015). On a more global level, speech rate is low (Evans & Deliyski, 2007; Lane et al., 1998; Leder et al., 1987; Perrin et al., 1999) but increases with implant experience (Oster, 1987; Perkell et al., 1992). Correspondingly, the duration of speech elements is long at different linguistic levels, such as syllables (Lane, Matthies, Perkell, Vick, & Zandipour, 2001; Menard et al., 2007; Neumeyer, Harrington, & Draxler, 2010; Uchanski & Geers, 2003), words (Kishon-Rabin, Taitelbaum, Tobin,

& Hildesheimer, 1999; Uchanski & Geers, 2003; Waters, 1986), sentences (Leder et al., 1987; Uchanski & Geers, 2003), and paragraphs (Leder et al., 1987). Perceptually, the voice of CI users is rated to some degree as strained, rough, breathy, asthenic, unstable and hoarse (Baudonck, D'Haeseleer, Dhooge, & van Lierde, 2011; Horga & Liker, 2006; van Lierde et al., 2005).

It could be argued that even within the population of CI users differences in hearing history have differential effects on voice and speech measures. For instance, postlingually deafened adults might benefit from feedforward articulatory commands established during the period as hearing individuals, whereas speakers with prelingual hearing loss or children with postlingual hearing loss had no or little opportunity to establish those commands (Perkell et al., 1992; Perkell et al., 1997). However, speaker groups with different onsets of hearing loss have been rarely tested in a single study. Hassan et al. (2011b) found higher nasality values relative to a NH control group for adults with more than six years of hearing loss than for adults with less than three years of hearing loss. Richardson, Busby, Blamey, Dowell, and Clark (1993) measured vowel formants in two adults and three children, but the sample size was too small to draw firm conclusions. The question to what extent voice and speech measures differ between adult and pediatric CI recipients therefore largely remains an open question. The current study focused on children.

Despite its broad range, the research on CI speech has failed to fully consider a number of important theoretical and methodological aspects. First of all, some prosodic measures have not been investigated phonetically, such as the natural declination of F0 during an utterance or the ratio of voiced and unvoiced frames. These specific measures are potentially interesting because they could reflect CI recipients' difficulty with perceiving F0. Second, basic measures of prosody, i.e., prosodic measures that have not been linked to a linguistic or emotional function, have, to our knowledge, not been systematically compared across phonetic dimensions within a single study. A comparison between the temporal, intensity, and spectral dimensions may allow connecting problematic phonetic aspects to auditory resolutions along those same dimensions. O'Halpin (2009) investigated accuracy of perception and production of duration, intensity and F0 cues of focused words, but this involved only one measure per dimension and was performed on laboratory instead of spontaneous speech. Third, measures were usually not compared at several points in time before and/or after implantation and/or for children with different ages at implantation. And finally, spontaneous speech has been neglected, even though voice differences can be expected between spontaneous speech and task-related speech (Vorperian & Kent, 2007). The use of spontaneous speech is important because it is the natural daily speaking mode. For instance, it could be argued that asking CI recipients to describe a picture, as in Evans and Deliyski (Evans & Deliyski, 2007), elicits a type of speech that is only spontaneous to a limited degree since the recipient is confronted not only with a specific semantic register but also with an experimental setting.

The present study aims to complement the body of research on CI users' speech characteristics by comparing a number of basic prosodic characteristics along three different phonetic dimensions in the spontaneous speech of young children: 'temporal', 'intensity', and 'spectral'. These dimensions were selected to reflect three important phonetic and acoustic parameters for which CI users have been found to have differential auditory resolutions and effectiveness (Cooper, Tobey, & Loizou, 2008; Moore, 2003; Shannon, 2002). This allows us to investigate to what extent perceptual competences are reflected in speech production. Measurements were repeated at three points in time after the onset of hearing and compared between children implanted before, or after the age of two years and a control group of normally hearing (NH) children of the same hearing age (Boons et al., 2012; Hayes, Geers, Treiman, & Moog, 2009; Holt & Svirsky, 2008). We conjectured that (1) the CI recipients' measures differed from those of the controls because they had less successful auditory feedback to control their laryngeal and articulatory output; (2) CI recipients were least deviant on the temporal dimension, followed by the amplitude dimension and most deviant on the spectral dimension; (3) the late implanted group had more deviant outcomes than the early implanted group; and (4) that the differences between CI recipients and controls decreased with increasing experience with the device and that this decrease was faster for early implanted than for late implanted children.

2.2 Methods

2.2.1 Participants

The study included three groups. There were two experimental groups, consisting of nine children implanted before and nine after the age of two, respectively (Early/Late Implanted, EI/LI; both 6 boys and 3 girls) with mean chronological ages of two years and ten months (henceforth, ‘2;10’; SD: 0;7) and 6;8 (SD: 2;5) at the time of testing. These participants were profoundly deaf and received a CI at Leiden University Medical Center (LUMC). The third (control) group consisted of 12 normally hearing children (4 boys, 8 girls) with a mean age of 2;1 (SD: 0;4; NH group). Eleven of them were children of the CLPF (Clara Levelt – Paula Fikkert) corpus (Fikkert, 1994; Levelt, 1994), available through the CHILDES database (MacWhinney, 2000) and through personal communication. One was from a corpus compiled by Beers (1995).

Demographic, audiometric and implant characteristics for individual CI recipients and for groups, as well as results of one-way Analyses of Variance of group mean differences can be found in Table 1. Some variables require an explanation. Age at onset of hearing loss diagnosis reports the age at which hearing loss was first diagnosed, with *0* for presumed congenital deafness. The estimated duration of deafness is the time between the estimated onset of deafness and age at CI activation. The mean age over recordings is the arithmetic mean chronological age of all recordings of a recipient that were used for analysis. This statistic was preferred over the age at first recording because not all sessions were available for all CI recipients (see the Data analysis section).

Groups were matched for hearing age, which is defined as the time since the onset of stable spoken language acquisition, i.e., without a changing hearing situation. For the CI group, this equals the time between CI activation and the time of recording; for the NH group, this equals the time between birth and the time of recording (i.e., chronological age). Matching for hearing age is a common procedure in CI language acquisition research, as language development of children with CIs has been found to match the development of NH children better by hearing age than by chronological age (Dornan, Hickson, Murdoch, & Houston, 2009; Fagan & Pisoni, 2010). This suggests that spoken language development starts with the onset of hearing and not necessarily at birth. Since in our study we were not interested in language development in general, but in phonetic development, we kept the amount of experience with stable spoken language input (i.e., hearing age) constant across participant groups.

Inclusion criteria for CI recipients were pediatric chronological age (under 11 years), bilateral pre- or postlingual severe-to-profound hearing loss, and a monolingual Dutch home environment. Exclusion criteria were reported additional social, cognitive or physiological disorders. All CI recipients were enrolled in the LUMC rehabilitation program for pediatric CI recipients, involving frequent speech training and six-monthly communication and social behavior follow-ups. The dividing line between Early and Late age of implantation was set at two years because differences in language outcomes have been observed between children implanted before or after this age, likely due to a boundary of one of the sensitive periods of language acquisition (Boons et al., 2012; Hayes et al., 2009; Holt & Svirsky, 2008; Werker & Hensch, 2015).

Matching groups for hearing age, combined with the selection by differential activation ages for different recipient groups

Table 1. Demographic and implant characteristics of CI recipients and the mean age of the control group. 'AB' is the Advanced Bionics HiRes 90k implant; 'Nucleus' is the Nucleus Freedom Contour Advance implant. BERA thresholds refer to the highest loudness levels in the left (L) and right (R) ear, respectively, that no BERA response was reported for. The group CI is the Early and Late Implanted groups taken together. SDs were rounded to whole months. Note that the (chronological) age and the hearing age are, by definition, the same for the NH group. Abbreviations: x;y.z – years;months.days. Numbers in parentheses indicate standard deviations, unless indicated otherwise. For Mean age over recordings and Mean hearing age over recordings, 2-way comparisons are Bonferroni corrected post-hoc analyses.

Group	Subject number (gender)	Age at onset of hearing loss diagnosis (months)	Estimated duration of deafness (months)	Age at CI activation	Mean age over recordings	Mean hearing age over recordings
	1 (M)	3	12	1;2.24	2;8.24	2;0.22
	2 (M)	0	13	1;1.20	2;8.28	2;1.18
	3 (M)	0	17	1;4.26	2;7.15	2;0.24
	4 (M)	0	12	0;11.26	2;7.08	2;1.26
EI	5 (F)	4	15	1;7.09	3;2.16	2;3.29
	6 (F)	2	16	1;5.23	3;1.28	1;10.7
	7 (M)	1	13	1;2.00	2;7.19	1;5.20
	8 (F)	4	10	1;1.26	2;6.23	1;8.15
	9 (M)	7	11	1;6.12	3;0.08	1;11.29
	MEAN	***2.3 (2.4)***	***13.2 (2.3)***	***1;3.19 (0;2.16)***	***2;10.9 (0;6.18)***	***1;11.18 (0;3.4)***
	1 (M)	0	49	4;1.08	5;4.05	1;10.12
	2 (F)	16	27	3;6.23	5;3.04	2;1.1
	3 (F)	30	16	3;9.17	5;3.04	2;0.18
	4 (M)	0	96	8;0.00	9;6.28	2;1.1
LI	5 (M)	16	86	8;5.28	10;2.02	2;0.24
	6 (M)	9	64	6;0.19	7;6.16	2;0.1
	7 (M)	12	47	4;10.22	6;4.08	1;5.20
	8 (M)	2	81	6;10.16	8;4.27	1;10.11
	9 (F)	0	25	2;1.27	3;7.18	2;0.7
	MEAN	***9.4 (10.2)***	***54.6 (28.9)***	***5;3.28 (2;1.27)***	***6;8.12 (2;4.22)***	***1;11.18 (0;2.12)***
CI	***OVERALL***	***5.9 (8.1)***	***33.9 (29.1)***	***3;3.23 (2;6.18)***	***4;9.11 (2;7.4)***	***1;11.13 (0;2.22)***
NH	***MEAN***				***2;0.15 (0;3.29)***	***2;0.15 (0;3.29)***
3-way	ANOVA *p* (*F*)				**<.001** (32.9)	.69 (.37)
EI-LI	ANOVA *p* (*F*)	0.059 (4.1)	**.001** (18.0)	**<.001** (31.0)	**<.001**	1
EI-NH	ANOVA *p* (*F*)				.54	1
LI-NH	ANOVA *p* (*F*)				**<.001**	1
CI-NH	ANOVA *p* (*F*)				**.002** (11.8)	.39 (.77)

Notes: [a] Calculations were based on available cases and on means of both ears where applicable

Table 1 (cont.)

Group	**Subject number (gender)**	**Etiology**	**BERA threshold L/R (dB)**	**Implan-ted ear(s)**	**Implant type**	**Speech process-sor**	**Insertio dept (degrees**
	1 (M)	unknown	92/90	bilateral	AB	PSP	467.99/483.1
	2 (M)	hereditary	95/100	right	AB	PSP	480.4
	3 (M)	unknown	108/103	right	AB	PSP	461.3
	4 (M)	hereditary	unknown	bilateral	AB	PSP	405.16/447.7
EI	5 (F)	unknown	103/103	bilateral	AB	PSP	465.53/425.1
	6 (F)	unknown	100/100	right	AB	PSP	547.7
	7 (M)	unknown	100/100	bilateral	AB	PSP	455.03/506.9
	8 (F)	unknown	105/105	right	AB	PSP	498.5
	9 (M)	unknown	100/100	bilateral	AB	PSP	437.05/560.5
	MEAN		***100.3 (4.6)***[a]				***479.47 (34.86)***
	1 (M)	unknown	100/100	left	AB	PSP	482.6
	2 (F)	meningitis	90/100	left	AB	Auria	575.6
	3 (F)	unknown	97/97	right	AB	Harmony	504.9
	4 (M)	unknown	100/85	left	AB	Harmony	
LI	5 (M)	unknown	90/90	left	Nucleus	Freedom	
	6 (M)	unknown	no response[b]	left	AB	PSP	
	7 (M)	unknown	100/80	left	AB	PSP	463.5
	8 (M)	meningitis	100/100	left	AB	PSP	512.9
	9 (F)	unknown	97/97	right	AB	Harmony	632.4
	MEAN		***95.2 (4.0)***[a]				***528.69 (63.46)***
CI NH	***OVERALL MEAN***		***97.7 (4.9)***				***499.16 (52.4***
3-way	ANOVA *p* (*F*)		**0.035** (5.42)				
EI-LI	ANOVA *p* (*F*)						0.073 (3.8

Note : [b] BERA performed in another medical center

unavoidably introduced a confound with chronological age. As can be seen in Table 1, therefore, measures relating to chronological age were statistically different between groups (except for EI vs. NH for chronological age), but not those relating to hearing age. The Spearman rank correlation between Group and Chronological age was 0.922. When fitting both Group and Chronological age into the statistical model (multilevel linear regression model), standard errors were highly inflated and parameter estimation became highly unstable. We therefore only considered the variable Group in the statistical model, without chronological age. We will return to this complication in the Discussion section.

EI recipients were implanted in the right ear (N = 4) or bilaterally (N = 5), whereas 7 out of 9 of the LI recipients were implanted in the left ear. All but one recipient received the Advanced Bionics HiRes 90k with a HiFocus 1j electrode and a PSP (including all the EI recipients), an Auria or a Harmony speech processor (Advanced Bionics, Sylmar, CA, USA); one recipient in the LI group was fitted with the Nucleus Freedom Contour Advance (Cochlear Corp, Sydney, Australia). Etiologies were unknown in most cases, except for hereditary causes and meningitis in two cases each. Insertion depth in degrees (computed as the mean between both ears if applicable) was not different between groups, but Brainstem Evoked Response Audiometric (BERA) thresholds were higher for EI than for LI.

2.2.2 Procedure

Speech recordings of the experimental participants were performed in playrooms at the department of pediatrics in LUMC. The setup consisted of a table, chairs, games and toys (such as cars and a kitchen) for children. A researcher observed and videotaped the session. Audio was recorded through the camera's integrated high-quality microphone or one attached to children and parents' clothing just below the head. Both in the recordings of the experimental and those of the control group, the child played with (a) parent(s) or a therapist/experimenter and sometimes also siblings. A child's speech was elicited when he/she did not speak much spontaneously. A recording session typically lasted between 20 and 30 minutes.

2.2.3 Data analysis

Audio channels were digitized with a 16-bit resolution and at a 48 kHz sampling frequency. Speech segmentation and phonetic analyses were performed by a trained linguist and phonetician (DV) using *Praat*

Table 2. List of prosodic measures performed for the analysis of the speech data, each listed under the phonetic dimension (temporal, intensity, spectral) that it is classified under for the current purpose. Abbreviation is the code by which it is referred to in the text (if unspecified, the full name is used). Unit is the mathematical unit used to describe an outcome of the measure. σ stands for syllable. Definitions are explained in the text.

Dimension	Measure (abbreviation)	Definition	Unit
Temporal	Articulation rate (ArtRate)	Number of syllables pronounced per second speech without pauses	σ/s
	Duration of the utterance (log) (DurUtt)	Base-*e* logarithm of the difference between final and initial time point of the utterance	s
	Voicing Ratio	Portion of voiced frames of an utterance as a percentage of the total number of analysis frames in the utterance	%
Intensity	Amplitude Perturbation Quotient (APQ)	(5-point scale). "The average absolute difference between the amplitude of a period and the average of the amplitude of its and its four closest neighbors, divided by the average amplitude."	%
	Harmonics-to-Noise Ratio (HNR)	The ratio between the energy that is in the periodic part and the energy that is in the aperiodic part of the voiced stretches of the signal	dB
Spectral	Declination	Global trend of F0 from beginning to the end of an utterance	Hz/s
	Mean F0	Mean of all pitch points (i.e., F0) of an utterance	Hz
	F0 standard deviation (SD F0)	Standard deviation of the mean of all pitch points (i.e., F0) of an utterance	Hz
	Pitch Perturbation Quotient (PPQ)	(5-point scale). "The average absolute difference between a period and the average of its and its four closest neighbors, divided by the average period."	%

software, *Version 5* (Boersma & Weenink, 2014). NH and CI recordings were matched for hearing age with a five-day margin per session (18, 24, 30 months). This yielded twenty recordings per group divided over hearing age sessions at 18, 24 and 30 months. Due to restricted data availability at source in combination with the strict matching criteria, this design suffered from missing data (see the section Statistical Analysis). All recordings were subjected to the same data processing procedure. Nine phonetic prosody parameters were measured (Table 2). We will call them ‘basic’ measures because they do not involve linguistic or subjective judgements about the (un)naturalness, function or meaning of the prosody. They cover three fundamental acoustic dimensions of prosody: the temporal, the intensity and the spectral dimensions (Lehiste, 1970). The temporal measures were articulation rate (ArtRate), duration of the utterance (DurUtt) and Voicing Ratio. ArtRate is defined as the number of syllables pronounced per second speech without pauses (Goldman-Eisler, 1968). Numbers of syllables per utterance were determined from the recordings, on the basis of the realized, not the targeted, form of words. The duration of the utterance (DurUtt) was based on prosodic and syntactic integrity. The exact starting and end points were based on visual inspection of the waveform. Voicing Ratio refers to the percentage of frames of an utterance that are voiced. This was based on a pitch analysis whereby the time-step for frames was 75 ms and the pitch range of analysis was 100-600 Hz. The reason we consider this a temporal measure is that correct production of voicing specifically requires that the timing of the onsetand offset of vocal fold vibration is synchronized with the sequence of vowels and consonants.

The intensity measures are the five-point amplitude perturbation quotient (APQ) and Harmonics-to-Noise Ratio (HNR). APQ is “[t]he average absolute difference between the amplitude of a period and the average of the amplitude of it and its four closest neighbors, divided by the average amplitude.”[1] This is a measure of local variability of the amplitude of an F0 period. HNR represents the ratio (expressed in dB) between the energy in the harmonics vs. the energy in the parts between the harmonics of the voiced stretches of the signal. Periodicity was detected using the cross-correlation method with a time-step of 10 ms, a pitch floor of 100 Hz, a silence threshold of 0.1 times the global maximum amplitude and 1 period per time window.[2] Despite the fact that HNR carries both spectral (absence or presence of periodicity) and intensity-related signal information, we regard the intensity-related information as primary, since HNR is defined as a ratio of intensities, and is therefore an intensity measure itself. These intensity measures could count as prosodic measures because they involve voice quality measured over a full utterance.

The spectral measures are declination of F0, standard deviation of F0, the mean of F0 and the pitch perturbation quotient. Declination is the natural global downtrend of F0 from beginning to the end of an utterance (Strik, 1994). To our knowledge, declination has never been estimated in CI users' speech. Because its realization depends not only on physiological effort but also on linguistic choices for which good control of F0 is needed, we expect that CI recipients will relatively often disrupt the baseline deviation such that values will become less negative (shallower downtrends). Mean F0 was calculated as the mean of all pitch points (i.e., F0) of an utterance. Following previous research, we expect to find elevated values of mean F0 for CI users (Oster, 1987; Perkell et al., 1992; Szyfter et al., 1996; Ubrig et al., 2011). The standard deviation of F0 (SD F0) is computed as the deviation of the mean of all pitch points of an utterance. It could be taken as a proxy for the global variability of F0 over an utterance. Based on research on a comparable measure, vF0, the coefficient of long-term F0 variation (the relative standard deviation of the period-to-period F0) (Deliyski, 1993; Hocevar-Boltezar et al., 2006; Holler et al., 2010; Ubrig et al., 2011), we hypothesize higher values for the CI recipients than for the controls. Finally, the five-point PPQ is "[t]he average absolute difference between a period and the average of its and its four closest neighbors, divided by the average period."[3] This is a measure of local pitch variability.

The utterance was used as the unit of the measurements, as this counts as a unit for many aspects of prosody. It is the highest prosodic unit under discourse-level units where intonational boundaries and temporal organization coincide (Rietveld & van Heuven, 2016). Utterances that were inaudible and/or interrupted by other speakers were left out because their phonetic realization and/or analysis would be unreliable. This yielded 1,973 utterances. From this set, in order to avoid improbable values due to pitch detection errors, utterances were removed from the analysis if the declination was more than two standard deviations away from the mean (1.8%), resulting in 1,937 utterances for analysis. Different participants provided different raw and net numbers of utterances, but all measures were performed for every available utterance.

A risk of using a corpus of spontaneous speech is that the speech material is not equal between groups. It is especially important for Voicing Ratio and, to a lesser extent, for ArtRate that the realized segmental material be phonetically balanced. We therefore obtained an approximation of the number of tokens per phoneme used in the whole data set of each Group. Figure 1 displays the token occurrence per phoneme as a percentage of the total number of tokens in the group. The graph shows that the distributions of allophone tokens are highly comparable between groups. A second possible pitfall in corpus research is the number of syllables. However, according to an ANOVA, there was no effect of Group on the mean number of syllables per participant ($F(2,27) = 1.25$, $p = .30$).

2.2.4 Statistical analysis

Statistical tests were carried out using *IBM SPSS Statistics, Version 21* (IBM, Armonk, NY). Each participant was measured at three planned occasions and each occasion provided multiple (unique) utterances. The statistical model took into account that utterances were correlated within participants. For each of the seven dependent variables separately, a multilevel linear regression model was used to describe the differences between the groups and between time points of measurement, with within-subject correlation being modelled by introducing a random subject intercept. This was done by modelling the correlation structure before the fixed structure (Fizmaurice, Laird, & Ware, 2011). The procedure started by applying a very complex and well-fitting model and subsequently reduced it using Restricted Maximum Likelihood and Maximum Likelihood Ratio tests. When a

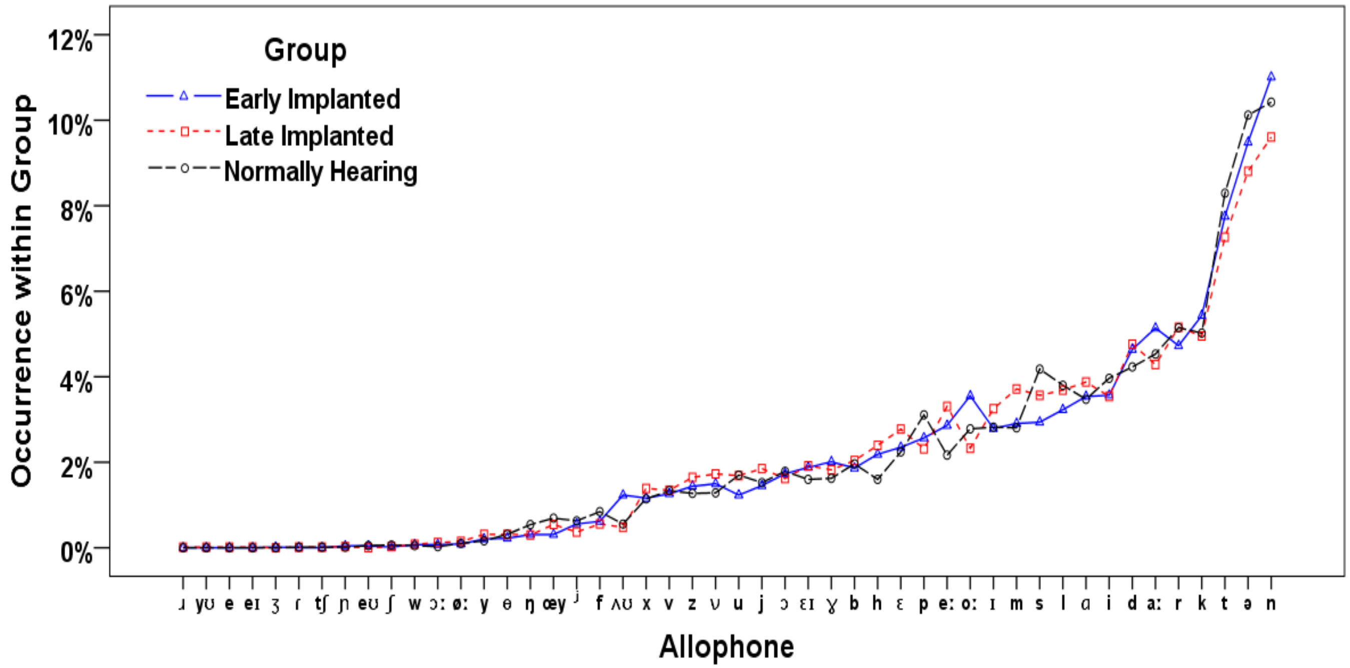

Figure 1. Token frequencies of 48 vowel and consonant phonemes of Dutch (N = 22,106) for three groups of speakers (Early Implanted, Late Implanted, Normally Hearing), as a percentage of all tokens within each respective group. Allophone, on the abscissa, refers to the allophone types found in the combined speech material of all participants. The transcriptions into these allophones were were left unaltered as they were created *Praat*. Type and token frequencies were obtained by automatically converting an orthographic transcription of all utterances to a broad phonetic transcription.

decision could not be based clearly solely on Likelihood Ratio tests, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were considered to decide on the most appropriate model (Fox, 2008). Models were fit using the Linear Mixed Model procedure in *SPSS*. A significance threshold of $p = 0.05$ was adopted.
In order to explore possible correlations among the nine dependent variables obtained for the analysis (see Table 2), an exploratory factor analysis using a principal component extraction method and a varimax rotation was conducted using heuristics and steps taken from Meyers, Gamst, and Guarino (2006). All correlation coefficients are shown in the correlation matrix in Table 3. The data were screened by considering both univariate and multivariate descriptive measures. All variables were interval variables and, except for DurUtt, approximately normally distributed. DurUtt was logarithmically transformed (with base *e*). Using these variables, all variable pairs appeared to be bivariate normally distributed with the exception of the pair ArtRate - DurUtt. The Kaiser-Meyer-Olkin measure of sampling adequacy for this pair was 0.612, which is not considered adequate given a criterion of 0.7. However, a factor analysis showed that three variables were correlated to a medium to high degree, viz. HNR, PPQ and APQ. Considering only these three relatively strongly correlated variables, the Kaiser-Meyer-Olkin measure was adequate (0.707). Bartlett’s test of sphericity was, however, significant both when including and excluding the three non-highly correlated variables ($\chi^2(36) = 4032.65$, $p < .001$; $\chi^2(36) = 2919.03$, $p < .001$). We concluded that the dataset was appropriate for factor analysis. In the factor analysis considering all nine dependent variables, four eigenvalues greater than 1 were found (2.553, 1.404, 1.078, and 1.044).

Given the preference for interpretable dependent variables, and also taking into consideration that the second principal component consisted of two variables with only a small correlation (0.280), only the first component was constructed. The factor (henceforth, Factor 1) was constructed by standardizing and summating the three dependent variables that were involved in the component (HNR, PPQ and APQ). Further analysis was thus done using the seven (almost uncorrelated) dependent variables.

Table 3. Correlation matrix with coefficients of the Pearson correlations between the nine dependent variables, plus two-tailed significance indications and *p-value* (between parentheses). Definitions of the measures can be found in Table 2.

Measure	HNR	PPQ	APQ	Mean F0	SD F0	Voicing Ratio	DurUtt ($\log_e$)	ArtRate
PPQ	**−.598**							
APQ	**−.763**	**.674**						
Mean F0		**−.146**	**−.263**					
SD F0	**−.112**	**.111**	0.028	**.280**				
Voicing Ratio	**.274**	−0.044	**.045**	0.008	**−.106**			
DurUtt ($\log_e$)	**.118**	**−.128**	**−.174**	0.026	**.201**	**−.111**		
ArtRate	**.090**	**−.177**	**−.101**	**.090**	**.048**	0.034	**.163**	
Declination	−0.011	**.050**	0.037	**.049**	0.021	0.006	0.013	0.038

Notes: Correlations in boldface were significant. In this table, correlation coefficients >.045 were significant at the $p < .05$ level, and correlation coefficients >.090 were significant at the $p < .01$ level.

As explained in the section Data analysis, recordings were missing on one or two sessions for some participants. There were a number of causes: 1) the recording contained no or hardly any analyzable child utterances (1 case, EI); 2) the recording did not exist because the child had been implanted too recently (3 cases, EI); 3) the recording at that session was not performed because that was not deemed necessary by the speech therapist given his/her development or because some other test was performed during that visit (2 cases, LI); 4) technical problems (2 cases, LI); 5) the session fell outside the range ever recorded by an LUMC speech therapist for a participant (16 cases, NH). Recording selections were based on the chronological age during recording and not on the quality of their content. We therefore believe our data are Missing Completely At Random or perhaps Missing At Random (Fizmaurice et al., 2011) which allowed us to use a linear mixed model that uses the likelihood function to estimate the parameters in an unbiased way. For a recent review on the problem of and solutions for missing data in otorhinolaryngological research, see Netten et al. (2016).

In sum, seven independent linear mixed model (LMM) analyses were run, each for one of the dependent variables (one of which, Factor 1, is a combination of three of the original variables). We were interested in the effect of the independent variables Group (EI, LI or NH) and Session (a hearing age of 18, 24 or 30 months). Though its effect was not a focus in itself, the variable Gender of the participant was added as well, viz. in order to account for a possible confounding effect because genders were not equally divided across groups (see Table 1).

2.3 Results

Mean values and standard deviations (in parentheses) of all nine dependent variables and Factor 1 are listed in Table 4. This includes the values aggregated over one, two, and three independent variables (Group, Session and Gender). APQ, HNR, and PPQ will not be discussed separately, as they have been merged into Factor 1. Means and confidence intervals of the seven dependent variables left after factor analysis are shown in Figure 2. The development in hearing age in months (Session) was plotted on the abscissa. This was split by Group and Gender (left panels), and separately, for clarity, split by only Group (right panels).

The grouping of APQ, HNR, and PPQ into Factor 1 eliminated one of the phonetic dimensions under investigation, viz. the intensity dimension, as the two intensity measures were both part of that procedure. Results of the remaining seven variables will now be discussed in turn. Following the Principle of Marginality, main effects were not interpreted when more complex terms present in the model were significant (Fox, 2008). Further, individual regression coefficients were not interpreTable 1n those cases either, because they cannot be considered separately from the interactions. Table 5 lists the best-fit models and statistics of the component effects for all seven dependent variables. Best-fit models refer to the combination of terms

Table 4. Mean values and standard deviations (right sides of columns) of all nine dependent measures and Factor 1, divided over Group (EI: Early Implanted, LI: Late Implanted, NH: Normally Hearing), Gender, and Session (hearing ages of 18, 24, and 30 months). Factor 1 is the sum of z-transformed values of HNR, APQ, and PPQ. Definitions of the measures can be found in Table 2. ‘Syll’: syllable; ‘mos’: months.

Group	Ses-sion (mos.)	Measure											
		ArtRate (syll/s)		DurUtt (log_e, s)		Voicing Ratio (%)		APQ (%)		HNR (dB)		Declination (Hz/s)	
EI	18	2.27	.67	0.55	.09	0.68	.16	6.62	2.98	12.76	4.9	–8.16	101.33
	24	2.78	.77	0.58	.09	0.6	.14	4.62	1.55	14.3	3.62	–16.08	92.12
	30	2.86	.98	0.56	.11	0.64	.16	6.56	3.14	12.37	4.3	3.82	91.84
LI	18	2.94	1.21	0.51	.1	0.63	.18	7.55	4.23	10.47	6.11	–32.94	116.36
	24	3.3	1.1	0.54	.1	0.65	.17	6.04	2.97	13.13	4.51	–32.73	91.13
	30	2.78	.77	0.57	.13	0.64	.14	5.09	2.26	13.92	3.73	0.43	84.79
NH	18	2.22	.69	0.47	.05	0.75	.18	7.64	4.22	11.89	4.68	–56.57	110.78
	24	2.5	.81	0.52	.08	0.63	.15	5.69	2.14	13.38	3.79	–14.45	127.42
	30	2.78	.77	0.57	.09	0.62	.14	5.42	2.05	14.89	3.72	–4.7	66.52
Total	18	2.44	.83	0.51	.09	0.68	.18	7.25	3.85	11.7	5.37	–31.26	111.04
	24	2.7	.88	0.54	.09	0.63	.16	5.52	2.32	13.54	3.96	–19.34	111.77
	30	2.78	.85	0.57	.11	0.63	.14	5.66	2.51	13.96	4.03	–1.12	78.6
EI		2.63	.83	0.57	.1	0.63	.16	5.74	2.7	13.31	4.28	–8	94.76
LI		2.94	1.4	0.54	.11	0.64	.17	6.2	3.34	12.61	5	–24.08	98.05
NH		2.5	.75	0.53	.09	0.65	.16	5.86	2.58	13.65	4.01	–16.99	110.88
Total		2.63	.83	0.54	.1	0.64	.16	5.92	2.84	13.28	4.38	–16.43	103.48

Table 4 (cont.)

Group	Session (months)	Measure									
		Declination (Hz/s)		Mean F0 (Hz)		SD F0 (Hz)		PPQ (Hz)		Factor 1 (*z*)	
EI	18	–8.16	101.33	321.25	49.84	56.63	29.62	1.07	.53	0.35	3.01
	24	–16.08	92.12	325.16	53.92	61.11	24.51	0.97	.33	–0.89	1.82
	30	3.82	91.84	321.46	54.02	56.29	27.35	1.22	.54	0.94	3.02
LI	18	–32.94	116.36	310.73	63.24	53.95	31.49	1.33	.74	1.67	3.97
	24	–32.73	91.13	306.65	58.71	53.31	26.38	1.1	.47	0.13	2.77
	30	0.43	84.79	291.03	41.12	50.5	24.08	1.01	.37	–0.58	2.01
NH	18	–56.57	110.78	304	102.64	43.29	27.67	1.29	.65	1.35	3.47
	24	–14.45	127.42	330.08	48.2	51.83	23.04	1	.38	–0.28	1.95
	30	–4.7	66.52	304.46	33.49	48.17	21.93	0.98	.37	–0.74	2.05
	18	–31.26	111.04	312.42	73.69	51.72	30.19	1.23	.65	1.11	3.54
	24	–19.34	111.77	323.15	53.15	54.38	24.52	1.02	.4	–0.33	2.18
	30	–1.12	78.6	306	43.22	50.97	24.25	1.05	.44	–0.21	2.48
EI		–8	94.76	323.01	52.83	58.47	26.85	1.07	.47	–0.01	2.68
LI		–24.08	98.05	303.66	56.43	52.74	27.29	1.14	.55	0.37	3.1
NH		–16.99	110.88	318.72	56.44	49.57	23.53	1.03	.44	–0.21	2.33
Total		–16.43	103.48	315.9	55.96	52.83	25.74	1.07	.48	0	2.66

listed in the column Terms of the best-fit model in Table 5. Unless stated otherwise, the focus of the interpretation will be on Group and Session (the right panels of Figure 2), because Gender was considered a confounding variable. The left panels of Figure 2 are shown for the sake of completeness.

Table 5. Best-fit models and statistics of component effects for all seven measures left after factor analysis. The best-fit model refers to the combination of factors (Group, Gender, Session and all their interactions) that was found to be the best Linear Mixed Model for the data of each measure. It consists of the combined terms for that measure. See the text for the criteria used for finding the best-fit model. The statistics of component effects refer to the *F*-value, degrees of freedom and *p*-value found for each term in the best-fit model. df: degrees of freedom; significant differences (at $p = .05$) are in boldface. Degrees of freedom were rounded off to the nearest integer value.

Measure	Terms of the best-fit model	Statistics of the term			
		F	df1	df2	*p*
ArtRate	Group	1.97	2	24	.16
	Gender	6.42	4	186	**<.001**
	Session	10.05	2	217	**<.001**
	Group × Gender	2.11	2	24	.14
	Group × Session	6.60	2	217	**.002**
	Gender × Session	1.51	4	186	.20
	Group × Gender × Session	6.42	4	186	**<.001**
DurUtt	Group	.00	1	26	1.0
	Gender	.88	2	26	.43
	Session	57.23	2	1864	**<.001**
	Group × Session	12.16	4	1670	**<.001**
	Gender × Session	8.14	2	1780	**<.001**
	Group	.82	2	20	.45
	Gender	1.71	1	20	.21
	Session	7.55	2	182	**.001**
	Group × Gender	.48	2	20	.62
	Group × Session	7.82	4	156	**<.001**
	Gender × Session	5.34	2	182	**.006**
	Group × Gender × Session	2.05	4	156	.090
Decli-nation	Session	7.29	2	1402	**.001**
Mean F0	Group	.98		26	.39
	Gender	.094	1	26	.76
	Session	19.53	2	1897	**<.001**
	Group × Session	11.86	4	1880	**<.001**

SD F0	Group	4.95	2	23	**.016**
	Gender	.076	1	23	.79
	Session	5.76	2	1759	**.003**
	Group × Gender	2.44	2	23	.11
	Gender × Session	4.25	2	1759	**.014**
Factor 1	Group	.33	2	25	.72
	Gender	1.26	1	25	.27
	Session	30.11	2	1913	**<.001**
	Gender × Session	19.12	2	1888	**<.001**
	Group × Session	13.06	2	1828	**<.001**

The best-fit for ArtRate was with all separate (Group, Gender, Session) and combined independent variables together. Given that the three-way interaction is the most complex significant term, all other effects must be interpreted with caution. Articulation rates were on average 2.63 syllables/s (syll/s) for the EI group, 2.94 syll/s for the LI group, and 2.50 syll/s for the NH group. Panel 1b in Figure 2 shows that from 18 to 30 months, the EI and the NH children experienced a rise in ArtRate, with the EI being ahead of the LI, and that the LI children converged with NH starting from higher values. The EI were therefore closer to the NH than the LI on only one of the three sessions. To our knowledge, the only previous study comparing speech or articulation rates in children with and without CIs is by Perrin et al. (1999). They found lower rates for the clinical group than for the typically developing group. However, their participants were older (9 to 14 years) than ours and the researchers did not report absolute outcome values. The values of all groups in the current study were on the lower side but within the range reported in studies on 3- to 5-year-olds discussed in (Flipsen, 2002). Rates tended to increase with age (e.g., Amir & Grinfeld, 2011) and to be lower in atypically developing populations including (adult) CI users (Evans & Deliyski, 2007; Lane et al., 1998; Smith, Roberts, Smith, Locke, & Bennett, 2006). Recipients in the studies on CI were all implanted as adults. In the current study, groups were confounded by chronological age and groups with a higher mean age had faster rates. This suggests that pediatric cochlear implantation does not prevent the typical increase in articulation rate with age.

DurUtt was best fit with Group, Gender, Session, Gender × Session, and Group × Session. Interpretable are differences in development between Groups (our focus) and, separately, between Gender. Figure 2, Panel 2b shows that at 18 months the NH had the shortest utterances, the LI had longer utterances, and EI the longest, but there was a convergence over time towards high values, with LI showing a straighter development than EI. The LI, with 1.72 s (transformed back from the logarithmic value) were further away from the controls (1.70 s) than the EI were (1.77 s). Utterance or sentence lengths (measured in syllables, phones or seconds) of typically and atypically developing populations tended to increase with age (Flipsen, 2002 and references therein; however, see Kadi-Hanifi & Howell, 1992), but this was not currently reflected, as the oldest group (LI) did not show the longest duration. For our older participants (LI), the value was low in comparison to values mentioned in the literature. In one study on the unrestricted speech of three groups of 4-, 7-, and 11-yeard-old stutterers and age-matched non-stutterers (Kadi-Hanifi & Howell, 1992), the average durations of the first two control groups were both 5.15 s. This, together with the observation that values in the three groups of the current study, despite being significantly different, were in an absolute sense very close together, suggests that the utterance duration length depended not on the chronological age, but rather on the hearing age (which was matched between groups). The convergence over time could be due to differential mechanisms for the three groups, as suggested by a comparison between DurUtt and ArtRate. Because a higher articulation rate would, all else being equal, result in shorter utterances, the increase in DurUtt for the NH must be due to the number of syllables, the duration of silence within utterances, or both. To further investigate this possibility, mean numbers of syllables were computed (see the Data analysis section for the procedure) split between groups and sessions. For the 18, 24, and 30 months sessions, respectively, numbers of syllables were 2.2, 3.4, and 5.0 in the NH group, 3.7, 5.1, and 5.0 in the EI group, and 4.0, 5.0, and 5.3 in the LI group. According to an ANOVA, the interaction between Group and Session for this measure was highly significant ($F(4,1929) = 5.26$, $p < .001$). ArtRate and number of syllables per utterance developing more synchronously for the controls than for the CI recipients, it is very probable that control participants' utterances were longer because of an increasing number of syllables. The CI recipients, on the other hand, would tend to articulate faster on longer utterances without adding syllables. This could point at a more limited verbal working memory (compare, e.g., Burkholder & Pisoni, 2003). In conclusion, CI recipients' utterance duration seems to develop with hearing (not chronological) age and to be restricted by a relatively limited verbal working memory.

The best fit for Voicing Ratio was the one consisting of all separate and combined independent variables. The interpretable effects were Group × Session (this study's focus) and Gender × Session. In Figure 2, Panel 3b, it can be observed that CI recipients' Voicing Ratios started out lower than the controls' but converged towards comparable levels. The EI decreased in the first interval and were more variable, whereas the LI increased and were more constant. CI Recipients had a lower Ratio mainly at 18 months. EI children were not clearly more or less deviant than the LI children. It has been argued that children acquiring a first language pay attention to the distinction between voiced and voiceless intervals in the input in order to discover the rhythmic system of the language (Dellwo, Fourcin, & Abberton, 2007). Apparently, the implanted children did pay attention to this, but learned to time their voicing like NH peers 18 to 30 months after implantation.

The optimal fit for Declination was with only Session. Declinations became shallower over time, going from −31 to −1 Hz/s for all participants combined (Figure 2, Panel 4b; Table 4).

Declinations were less negative for the CI recipients, but mainly so at 18 months. EI participants were further from the NH values than LI at 18 months, closer at 24 months and about equally close at 30 months. These were only trends, however, since only the effect of Session was significant for Declination. ‘T Hart, Collier, and Cohen (2006) summarized the declination *D* of utterances under 5 s. in semitones per time unit as $D = -11/(t + 1.5)$, with *t* in seconds (also see Rietveld & van Heuven, 2016). This formula was found to both predict spontaneous and read-aloud utterances fairly accurately, although for spontaneous speech a somewhat shallower declination was reported. Given the overall mean F0 of 316 Hz and an overall utterance duration of 1.72 s. in our study, declinations of around −92 Hz/s were expected, which is much steeper than what we found (−16 Hz/s). This may be due to the fact that our participants were children, as it has been claimed that in very young children some units of speech (i.e., short ‘breath groups’) show no declination (Lieberman, 1986).

Mean F0 was best fit with Group, Gender, Session, and Group × Session. Mean F0 developed differently among Groups (Figure 2, Panel 5b). The EI children showed hardly any changes, whereas the LI children’s F0 dropped from 311 Hz at 18 months to 291 Hz at 30 months, and the NH children peaked in the middle session (from 304 to 330 Hz and back). With overall averages of 323, 304, and 319 Hz for the EI, LI, and NH groups, respectively. Mean F0 was, contrary to expectation, not higher in general in CI recipients, but only on two sessions for the EI and on one session for the LI. Further, EI were not clearly less deviant than the LI. The hypotheses regarding Mean F0 were therefore not confirmed. In one review of F0 values of children of different ages in 21 studies (Vorperian et al., 2005), the F0 value of one-and-a-half-year-old children (comparable to the mean age of the control group in the current study) was between 300 and 350 Hz, that of 3-year-old children (approximately the mean age of the Early Implanted group in the present study) ranged between 250 and 300 Hz and the value of the 7-year-old children (around the mean age of the Late Implanted group) ranged between around 240 and 280 Hz.

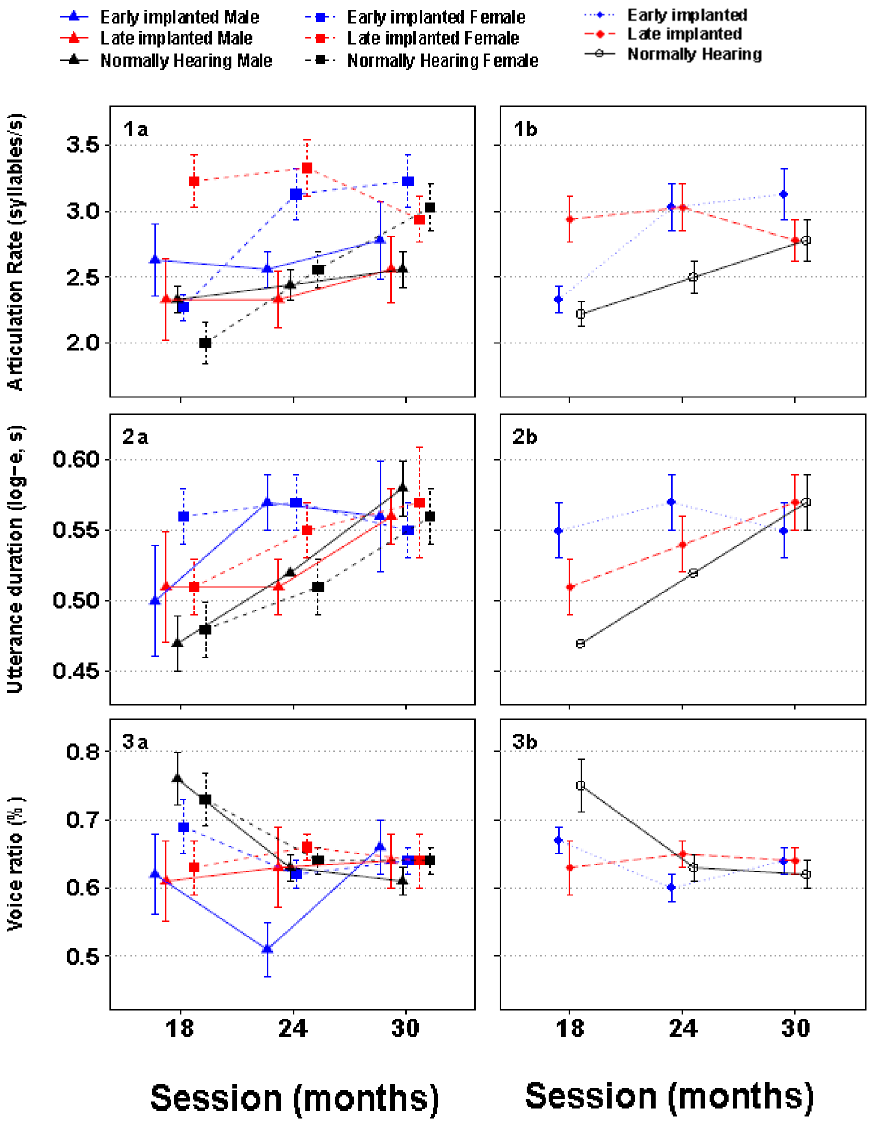

Figure 2. Plots of mean values of Articulation Rate, Utterance Duration (log-*e* transformed), Voicing Ratio, Declination, Mean F0, SD F0, and Factor 1. Factor 1 is the sum of *z*-scores of HNR, APQ, and PPQ. Hearing age in months (Session) is plotted on the abscissa. Left panels show results split by Gender, Group, and Session (Hearing Age in months). Right panels show the same results but aggregated over Gender. Error bars represent 95 % confidence intervals. The x-coordinates were jittered for the sake of clarity.

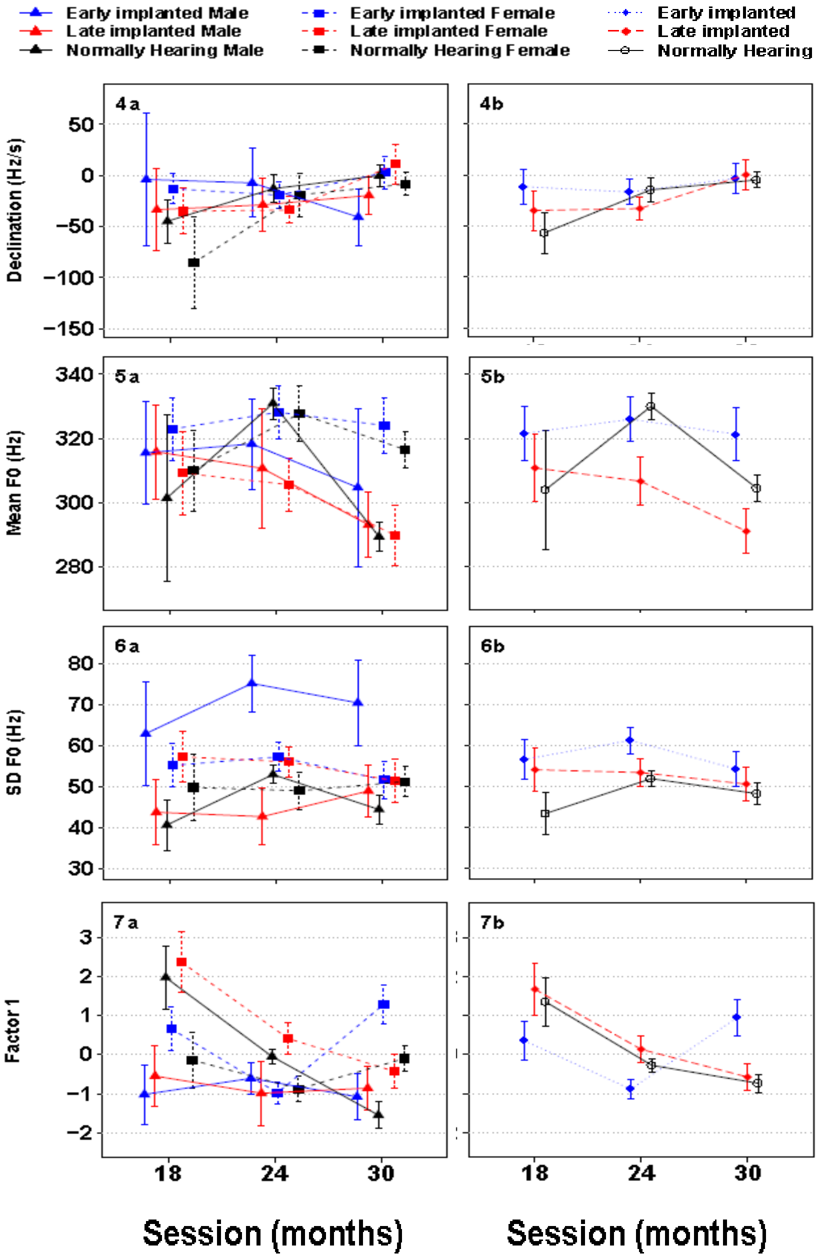

Figure 2 (cont.).

Interestingly, values of all our groups were in the range corresponding to the age of the youngest (NH) group, which suggest that hearing age, not chronological age, steered Mean F0.

SD F0 was best fit with Group, Gender, Session, Group × Gender, and Gender × Session. The Gender × Session interaction was the only interpretable effect. We can see in Figure 2, Panel 6a, that in general girls had extremer values and more variability than boys. There was, however, no overall difference in development between groups. The higher values for SD F0 for CI recipients (85.5 Hz for EI, 52.7 Hz for LI) as compared to controls (49.6 Hz) were in line with the predictions. The LI were, however, closer to the NH than the EI were. These values, especially those of the EI group, were considerably higher than those reported in an exploratory study on normative voice measurement values for younger and older adults (Goy, Fernandes, Pichora-Fuller, & van Lieshout, 2013), i.e., 26 Hz for males and 45 Hz for females. However, the participants in that study were much older (mean age 19.1 y. for the younger group) than those of the present study. This might explain the difference, as it has been suggested that with maturation children's voices become more stable (Kent, 1976). The literature shows mixed results concerning the effects of implantation age and implant experience on long-term frequency variability in implanted children. Holler et al. (2010) observed only an effect of time in sound (i.e., the sum of the time before the onset of deafness and the time since implant activation). Hsu et al. (2013) found an improvement (i.e., reduction of variability) as a function of experience, but no effect of implantation age. In a study by Campisi et al. (2005), there was no influence of implantation age nor of device experience. The current study is in agreement with results showing a convergence over time to normal values and more normal starting values for later implanted children.

Factor 1 was fit with Group, Gender, Session, Gender × Session, and Group × Session. Interpretable are the effects of Gender × Session and, our focus, Group × Session. Factor 1 was a combined factor. It therefore did not afford a prediction in the direction of possible deviation nor for a direct comparison with previous research. The high correlation of the three variables of Factor 1 (APQ, HNR and PPQ) is in agreement with previous literature (Hillenbrand, 1987). The measures most likely all stem from glottal pulse irregularity. Higher PPQ relates to higher APQ, in part because the energy from one pulse interacts with the energy from the next, more variability in pulse duration resulting in more variability in inter-pulse intensity resonance. The correlation between HNR and perturbation measures is due to shifts in measured zero-crossings (PPQ), and contributions to the pitch-pulse amplitudes (APQ) as a result of added random fluctuations, respectively (Hillenbrand, 1987). Because of this mechanism underlying the correlation between its three measures, we consider Factor 1 as the laryngeal factor. As reflected in Figure 2, Panel 7b, the LI children developed in parallel with the control group, following a downward trend, whereas the EI children had their very own trajectory, starting lower and ending higher. This could entail that laryngeal control requires maturation more than speech experience.

To summarize, we predicted that prosodic measures would differ between participant groups, with larger deviations from the norm for the LI than for the EI children. No interpretable main effects of Group were found, but we did observe a significant three-way interaction (Group × Gender × Session) on ArtRate as well as significant interactions between Group and Session, indicating differential developments, on DurUtt, Voicing Ratio, Mean F0, and Factor 1. For the Group × Session interactions, the LI showed a more constant development (or lack of development) than the EI on DurUtt, Voicing Ratio, and Factor 1, but not on Mean F0, where the EI were very constant but where the LI's values decreased much more. The LI's values were closer to the NH's than the EI's value on DurUtt, two out of three sessions of Mean F0, and Factor 1, but not on Voicing Ratio, where the two recipient groups were about equally different from the controls. On Declination and SD F0, no main effect of or interaction with Group surfaced as significant.

2.4 Discussion

The aim of this study was to compare the development of two dimensions of phonetic measures of prosody in the spontaneous speech of children with early (EI) and late (LI) cochlear implantation with those of normally hearing (NH) peers. These dimensions were the temporal (Articulation Rate, Utterance Duration, Voicing Ratio) and the spectral (Declination, Mean F0, Standard Deviation of F0) dimensions. A separate factor (Factor 1) was constructed as an arithmetic combination of Amplitude Perturbation Quotient, Harmonic-to-Noise Ratio and Pitch Perturbation Quotient. On both dimensions, deviations for CI recipients have been observed in the literature, but they have not systematically been compared in spontaneous speech production across different measures. We predicted that (1) CI recipients and controls would differ from each other, (2) they would differ least on the temporal and most on the spectral measures, (3) EI children would differ less from controls than LI children and (4) differences from the norm would diminish with increasing implant experience.

First of all, there were two confounding factors in this study, viz. chronological age and gender. We will discuss these two issues. As outlined in the Statistical Analysis section and Table 1 (see the column ‘Mean age over recordings’), the three participant groups had statistically different mean chronological ages. This was an unavoidable consequence of selecting for differential implantation ages while matching for hearing age. We have to take into consideration that any differences found between these groups could in principle also have been caused by age differences, or a combination of hearing age and chronological age. There are, however, two arguments to consider the age effect negligible. First, as an approximation of the effect of chronological age, we obtained Pearson correlations between chronological age and all of the dependent variables for all Group, separately. Out of 27 (i.e., 9 variables × 3 groups) cells, 13 correlations were below 0.1, 8 were below 0.2, while the largest coefficient was 0.409. This suggests that chronological age does not greatly influence any of the dependent variables. Second, for some measures, the pattern of results is not consonant with what would be predicted on the basis of the groups' chronological age. DurUtt is expected to increase with age, but the oldest group (LI) had values in between those of the other groups. On Voicing Ratio, groups did not clearly differ (apart from their developmental path). For Declination, the Group effect was not significant, but a trend (shallower declinations for older children) contrary to hypothesis could be discerned for two out of three Sessions. The values of Mean F0 are anticipated to drop with age, but a clear difference (i.e., independent of Session) in that direction was only observed between the two recipient groups and, moreover, that difference was smaller than what was suggested by the literature given the age difference between the groups. On SD F0, the oldest group (LI) was below the middle group (EI) but they were both above the youngest group (NH). For these reasons we conclude that the role of chronological age is small at most and does not prevent us from drawing conclusions based on differences between groups. When there are no differences between groups, it can be argued that results are dependent on hearing age, not chronological age. When the CI recipients' values are too low or too high relative to the age of the NH group, this is a sign that their hearing status influences the prosodic parameters of their voice. When the same pattern of results anticipated based on age is shown for all groups, this can be interpreted as a sign that cochlear implantation does not prevent a normal age-based development for this measure.

The second confounding factor was Gender. Gender was involved in effects on most measures (all but F0 and Declination) and, given that proportions of Gender were not equal across groups, that factor could potentially explain (some of) the effects of Groups. But note, first, that the proportion of Gender was only different between controls on the one hand and CI recipients on the other hand (i.e., not between the two recipient groups). And second, whereas girls were more variable in their development on DurUtt and Factor 1, the NH, despite their higher proportion of girls, were not more variable than the CI recipients. Likewise, the extremer and straighter development on Voicing Ratio and SD F0 for girls was not reflected in the trajectory of the NH group. We therefore feel safe to conclude that Gender is not responsible for differences in comparisons between recipient groups and the control group.

Our hypotheses were partly borne out. The first hypothesis (the CI recipients' measures differ from those of the controls) was supported for some, but not all, measures, although always in interaction with Gender and/or Session. This implies that hearing through a cochlear implant affects the development of speech due to the period(s) of atypical auditory sensations before and/or after implantation. This is in line with earlier literature reporting vocal deviations for CI children (e.g., Baudonck et al., 2015; Evans & Deliyski, 2007; Hocevar-Boltezar et al., 2006; Horga & Liker, 2006; Lane et al., 1998; Neumeyer et al., 2010; Oster, 1987; Poissant et al., 2006; Szyfter et al., 1996; Ubrig et al., 2011; van Lierde et al., 2005). This could imply that the atypical hearing situation of this population affects its vocal output in a general sense. It does not, however, specify to what level of perceptual detail this connection has an effect, i.e., if all acoustic parameters would be equally affected or if more problematic parameters would be more affected than relatively successful parameters. Our second hypothesis, the main focus of this study, was aimed at shedding light on that issue. We conjectured that CI users' voice deviances would be larger for the spectral measures, and smaller for temporal measures. This prediction was not in general supported by the results. The developments of Groups differed on three temporal measures (DurUtt, Voicing Ratio, and, in interaction with Gender, ArtRate), one spectral measure (Mean F0), and on the laryngeal factor (Factor 1). No effect was found for two spectral measures (Declination and SD F0). Importantly, this suggests that there is no clear correspondence between the degree of perceptual difficulty with a phonetic parameter and proficiency for that same parameter in production, as the poorer resolution for the spectral as opposed to the temporal dimension of the auditory signal was not reflected in a pattern of more deviant spectral than temporal speech measures.

Several previous studies have addressed the question of the relationship between perception and production performance of pediatric CI recipients. Peng and colleagues investigated Mandarin tone recognition and production by means of picture selection and naming, respectively (Peng, Tomblin, Cheung, Lin, & Wang, 2004). Across their thirty participants, they found a significant (r = .44) inter-test correlation. It has to be noted, however, that the correlation became non-significant when the top three performers were removed from the analysis. In another study, they compared appropriateness of elicited utterances' intonation with question vs. statement discrimination, finding a correlation of r = .65 (Peng, Tomblin, & Turner, 2008). Children with and without CIs in a set of experiments by O'Halpin (2009) had to decide whether utterances were compounds or phrases (e.g., *bluebottle* vs. *blue bottle*) and to identify which word in a phrase carried a focal accent. Scores on those tasks were compared to the participants' difference limens for F0, intensity and duration of synthetically manipulated nonsense syllables. O'Halpin concluded that the implanted children payed least attention to F0 cues, more to amplitude cues and most to duration cues. In production, however, these dimensions did not clearly differ from each other in their level of appropriateness. Moreover, interestingly, no correlations between participants' appropriateness of production and reliance on the acoustic dimensions was found except that an appropriate production of amplitude and duration was more related to a good perception of duration than of amplitude or F0. The results of this study suggest that despite differential perceptual competence of acoustic dimensions, this is not generally reflected in differential competence of those dimensions in production.

Nakata, Trehub, and Kanda (2012), testing Japanese pediatric CI recipients and NH controls, found a correlation of r = .56 for scores on prosody-based emotion recognition and rated appropriateness of imitated prosody. In a study on Mandarin-speaking children, Zhou, Huang, Chen, and Xu (2013) reported a significant correlation (r = .56) between accuracy for lexical tone identification on a picture selection task, and intelligibility of tones produced by picture naming. If broken up into individual tones, the correlation was significant for only two out of the four tones tested.

Taken together, studies about the perception and production of prosody in CI users, although not consistently, provide some evidence of a relationship in performance abilities between the two. There is, however, no evidence for a relationship per acoustic dimension, i.e., perceptual performance on a specific dimension does not predict the performance on that dimension in production. The present study is in agreement with the latter finding, since no clear advantage for a presumably better dimension (temporal over spectral) was observed. A number of explanations for the lack of correspondence between perception and production in the current study could be proposed. First of all, for speakers in general, the proficiencies in production and perception of speech could be independent of each other. This, however, appears not to be the case, given that the present study as well as previous work have demonstrated that there are discrepancies in the speech of individuals with hearing impairment with or without cochlear implants (e.g., Evans & Deliyski, 2007; Lane et al., 1998; Oster, 1987; Perkell et al., 1992; Perrin et al., 1999; Seifert et al., 2002; Szyfter et al., 1996; Ubrig et al., 2011) (Ball & Ison, 1984; Fourcin et al., 2011; Kishon-Rabin et al., 1999; Menard et al., 2007;

Nguyen et al., 2008; Svirsky et al., 1998). As a more direct indication, speech is altered soon after temporarily switching a CI off or back on (Higgins et al., 2001; Monini et al., 1997; Poissant et al., 2006; Svirsky et al., 1992; Tye-Murray et al., 1996). A second, more plausible account, therefore, would be that there is a relationship between production and perception, but that the difference in auditory resolution between the two dimensions currently studied is not large enough to result in a difference in production. This is also unlikely since the spectral and temporal resolution for most CI users cover two extremes, from very good to very poor, respectively (Moore, 2003; Shannon, 2002; Vorperian & Kent, 2007). A third possibility is that, although the spectral dimension is poorly processed, it is produced successfully because it is an automatic by-product of speech, i.e., it does not involve conscious linguistic or paralinguistic choices but is a physiological consequence of choices in other dimensions that may be consciously controlled. For instance, increasing a syllable's intensity for emphasis might be automatically paired with elevated pitch due to accelerated vocal fold vibration. Indeed, the two spectral measures showing a good performance, declination and SD F0, could be considered relatively uncontrollable variables, whereas the worse performance of Mean F0 could reflect its controllable nature. On the other hand, Factor 1 was relatively deviant, but would count as a less consciously controllable variable. Moreover, deviations in the temporal dimension would not be expected even for controllable variables, but they were found. All temporal measures were, however, in fact deviant as well as controllable and therefore it could be hypothesized that controllability plays a more important role than auditory resolution. This account is supported by at least two other considerations. First, our finding that CI recipients articulated faster on longer utterances (more so than the controls) could point to a limited verbal working memory span (Burkholder & Pisoni, 2003). That same limitation would also be part of the origin of a lack of control in the cases of prosodic parameters that require pronunciation choices assuming that would also be relatively taxing for verbal working memory. Second, the account would be in line with the claim that a lack of auditory feedback affects long-term parameters more than short-term parameters (Hsu et al., 2013), as both distinctions contrast the more linguistic with the more physiological parameters. Taking the above considerations together and abstracting away from underlying causes, we conclude that the quality, or lack thereof, of the acoustic speech dimensions received by implanted children is not directly reflected in comparable quality in those dimensions in their output, but that instead the controllability of prosodic voice parameters seems to be a more determining factor.

Our third hypothesis was that the LI would show more deviant outcomes than the EI group because they experienced a longer period without stable auditory input. LI's values were in general closer than the EI's to the NH's values, viz. on a temporal parameter (DurUtt), part of a spectral factor (Mean F0) and Factor 1, but not on another temporal measure (Voicing Ratio). Further, the LI children showed a less changeable development than the EI children on two temporal measures (DurUtt, Voicing Ratio) and the laryngeal factor, but it was the other way around for one spectral measure (Mean F0). Therefore, it seems that LI children did not deviate more than EI children; if anything, it was the other way around. This is in disagreement with most of the literature on the language development of CI users, where earlier implantation is associated with outcomes closer to the norm or with faster development. One possible cause for this is that four out of nine LI children had a late onset of hearing loss (between 12 and 30 months). This might have given them an advantage relative to the EI group, since in the time spent with relatively normal hearing prior to hearing loss they would have had some opportunity to establish speech goals from which they could still benefit after implantation. This could have partly compensated for the possible disadvantage from late implantation, resulting in less difference between the LI and EI groups.

Another possible cause is the fact that we focused on the more specific issue of voice and speech measures. Within the literature about age effects, few studies have done that. Advantages for earlier implantation or longer time in sound at various ages have been found regarding various segmental and suprasegmental variables (Tobey et al., 1991), glottal measures (Hocevar-Boltezar, Vatovec, Gros, & Zargi, 2005) and nasality (Hassan et al., 2011b), but not for formant values (Neumeyer et al., 2010). In one longitudinal study, prelingually deaf CI recipients showed a faster improvement but with more deviant starting values than postlingually deaf adults on a range of glottal measures (Hocevar-Boltezar et al., 2006). The results of the present study add to this overview by supporting the studies showing no benefit of earlier implantation (at any age) for prosody production. Instead, it does for some measures but not for others, possibly reflecting a compensatory combination of factors relating to perceptual resolution, controllability, implantation age and duration of hearing loss of the CI recipients. Future research should address a greater variety of measures and participant groups within a single study to disentangle these factors.

The fourth hypothesis stated that the differences between CI recipients and controls would decrease with increasing experience with the device and that this decrease would be faster for the early implanted than for the late implanted children. Groups converged over time on ArtRate (in interaction with Gender), DurUtt, Voicing Ratio, to some extent on Factor 1 (only LI and NH), and as a tendency on Declination and SD F0, but there was no convergence on Mean F0. These findings suggest that experience with the implant brought most voice parameters closer to the norm. This effect was stronger for temporal than for spectral measures. It held irrespective of implantation age. Our results resonate with previous reports showing improvement of some voice measures with increasing implant experience (Hassan et al., 2011b; Hocevar-Boltezar et al., 2006; Lenden & Flipsen, 2007), and especially research showing improvement of temporal (Goffman et al., 2002) but not spectral (Campisi et al., 2005) measures. Taken together, our results underline the suggestion that implant experience has a positive effect on prosody production, but more consistently so for temporal than for spectral measures.

Conclusions and future directions

The current study suggests that the appropriateness of different phonetic dimensions of the basic prosody of an utterance did not directly reflect the auditory resolution for the corresponding acoustic dimensions. The higher resolution for temporal structure than for spectral detail did not in general entail more successful production of temporal than spectral aspects of prosody in an utterance. Instead, it seemed that the parameters that required a relatively high level of articulatory and/or laryngeal control or planning (ArtRate, DurUtt, Voicing Ratio, Mean F0 and perhaps DurUtt) were somewhat more problematic than the parameters that were by-products of speaking (Declination, Factor 1, and SD F0). The data in this study did not shown an advantage of implantation before vs. after two years of age, but the outcomes improved with increasing implant experience.

The results of this study could be used as a recommendation for speech therapists to pay attention to the early development of basic prosodic measures of implanted children. I.e., using recordings of relatively spontaneous speech, they would have to monitor the measures that are at the risk of deviating and rehearse the necessary glottal and articulatory control and verbal working memory. It should be noted that the development of prosody can differ between parameters, between early and late implanted children and between genders. In future research, more different phonetic parameters should be compared in order to investigate more deeply the underlying cause of problems with some but not other parameters. It is also recommended that production results are directly compared with individuals' auditory resolutions on different dimensions, in an attempt to elucidate the possible correlation between perception and production in children with cochlear implants. Finally, in order to more clearly separate the effects of chronological age and hearing age, it would be advisable to orthogonally compare those two factors by testing early and late implanted children with the same chronological age, on the one hand, and with the same hearing age, on the other.

Acknowledgements

We are thankful to the Leiden Institute for Brain and Cognition (LIBC) for supporting this research. We are also grateful to ing. Jos Pacilly of the Leiden University phonetics laboratory for support involving signal processing. We thank Walter Verlaan of the LUMC, who helped with software and hardware issues related to recording and digitization. Statistician Vincent Buurman of the Faculty of Social Sciences helped us tremendously in analyzing the data.

[1] *Praat* manual, *Voice 3. Shimmer*.

[2] With these settings for analyzing HNR, analysis windows did not overlap, since with children's typical the analysis window is shorter than the time-step of 10 ms. With this procedure results are not based on the complete signal. In an informal comparison of the two procedures (non-overlapping vs. overlapping with 4.5 windows per period) the HNR values in the non-overlapping procedure were shown to be between 10% and 50% higher than with the overlapping method. It therefore has to be taken into account that with the overlapping method, lower HNR values would have been found.

[3] *Praat* manual, *Voice 3. Jitter*.