Perception of spectrally degraded reflexives and pronouns by children

Deniz Başkent[a),b)]
University of Groningen, University Medical Center Groningen, Department of Otorhinolaryngology/Head and Neck Surgery, P.O. Box 30.001, 9700 RB Groningen, The Netherlands

Jacolien van Rij
Eberhard Karls Universität Tübingen, Seminar für Sprachwissenschaft/Quantitative Linguistik, Wilhelmstraße 19, 72074 Tübingen, Germany

Zheng Yen Ng
University of Groningen, Center for Language and Cognition Groningen, P.O. Box 716, 9700 AS Groningen, The Netherlands

Rolien Free[b)]
University of Groningen, University Medical Center Groningen, Department of Otorhinolaryngology/Head and Neck Surgery, P.O. Box 30.001, 9700 RB Groningen, The Netherlands

Petra Hendriks
University of Groningen, Center for Language and Cognition Groningen, P.O. Box 716, 9700 AS Groningen, The Netherlands

(Received 20 December 2012; revised 18 September 2013; accepted 23 September 2013)

Speech perception skills in cochlear-implant users are often measured with simple speech materials. In children, it is crucial to fully characterize linguistic development, and this requires linguistically more meaningful materials. The authors propose using the comprehension of reflexives and pronouns, as these specific skills are acquired at different ages. According to the literature, normal-hearing children show adult-like comprehension of reflexives at age 5, while their comprehension of pronouns only reaches adult-like levels around age 10. To provide normative data, a group of younger children (5 to 8 yrs old), older children (10 and 11 yrs old), and adults were tested under conditions without or with spectral degradation, which simulated cochlear-implant speech transmission with four and eight channels. The results without degradation confirmed the different ages of acquisition of reflexives and pronouns. Adding spectral degradation reduced overall performance; however, it did not change the general pattern observed with non-degraded speech. This finding confirms that these linguistic milestones can also be measured with cochlear-implanted children, despite the reduced quality of sound transmission. Thus, the results of the study have implications for clinical practice, as they could contribute to setting realistic expectations and therapeutic goals for children who receive a cochlear implant.
 [http://dx.doi.org/10.1121/1.4824341]

PACS number(s): 43.71.Ky, 43.71.Sy, 43.66.Ts [EB] Pages: 3844–3852

I. INTRODUCTION

Cochlear implants (CIs) are prosthetic hearing devices that help people with profound to severe hearing impairment to partially regain the auditory function. Due to their proven success over the years, deaf children can now be implanted from very young ages, usually starting from 12 months and onwards (Geers and Nicholas, 2013). These young pediatric CI users greatly benefit from the implants (Schorr *et al.*, 2009), especially in terms of language development (Vermeulen *et al.*, 1999; Svirsky *et al.*, 2004; Nicholas and Geers, 2006). However, there seem to be limits to the sensitivity period; children implanted at around 3 yrs of age and younger seem to attain the best benefits (Kirk *et al.*, 2002; Sharma *et al.*, 2002), while the plasticity seems to drop significantly after the age of 7 yrs (Sharma *et al.*, 2002). Children implanted at older ages more often show deficits in speech and language performance compared to their age-matched normal-hearing counterparts. Considering the serious consequences such deficits can have for these children, it is crucial to fully characterize the language development and to accurately diagnose its potential delays due to hearing impairment, so that adequate rehabilitative solutions can be found. In clinical practice, typical methods used are questionnaires, such as Reynell's Developmental Language Scales (Reynell and Gruber, 1990), which characterize perceptive language development, and behavioral tests, such as speech perception tests with pre-recorded syllables, words,

[a)]Author to whom correspondence should be addressed. Electronic mail: d.baskent@umcg.nl
[b)]Also at: University of Groningen, Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands.

 0001-4966/2013/134(5)/3844/9/$30.00

or sentences (e.g., Nilsson *et al.*, 1996), which characterize speech intelligibility. While the questionnaires tend to be more linguistically based, due to their subjective nature they may not be sufficiently sensitive. On the other hand, while the behavioral tests are less subjective, they do not take into account important linguistic distinctions and may therefore fail to capture specific linguistic milestones.

In this study, we propose a new behavioral test for measuring language comprehension in children, using a linguistic task based on the comprehension of reflexives and pronouns (Van Rij *et al.*, 2010). In normal-hearing Dutch children, pronouns (e.g., *hem* "him") emerge in their spontaneous production between the age of 2 and 3 yrs old (Rozendaal, 2008). Reflexives (e.g., *zich* or *zichzelf* "himself") are much more infrequent than pronouns in children's production, as in the adult language input, but are produced in an adult-like way in elicitation experiments from the age of 4 yrs old on (Spenader *et al.*, 2009). In the present study we focus on the perception of reflexives and pronouns. Children's comprehension of reflexives is adult-like at age 5 (Chien and Wexler, 1990, with English-speaking children between the ages of 2 and 6; Van Rij *et al.*, 2010, with Dutch-speaking children between the ages of 4 and 6). However, adult-like comprehension of pronouns emerges later and does not show until the age of 10 (Koster, 1993; Philip and Coopmans, 1996). Note that in general children tend to make somewhat more errors than adults on linguistic tasks, as a result of their limited attention span and other interfering factors. Taking into account the errors caused by these non-linguistic factors, adult-like performance is usually understood to be a high percentage of correct responses, but not necessarily 100% (e.g., above 90%). The observation that adult-like comprehension of reflexives emerges early in childhood, whereas the adult-like comprehension of pronouns emerges much later is called the Delay of Principle B Effect (DPBE). The DPBE in normal language development has been firmly established for languages such as English and Dutch, although it is absent in various other languages (see, e.g., McKee, 1992, for Italian, and Ruigendijk *et al.*, 2010, for German). Because of the different developmental trajectories of reflexives and pronouns (in English and Dutch), a test based on the DPBE could be more sensitive to atypical or delayed language development than the standard clinical tests.

In the present study, as a first step, as well as to provide comprehensive baseline data on reflexive and pronoun perception in CI children, we explored the effect of spectral degradation of CI speech transmission using a noiseband vocoder simulation with normal-hearing children. In this simulation, slow-varying amplitude envelopes from a limited number of spectral channels are preserved, as would happen in an actual CI, while the fine structure in each channel is mostly removed by replacing it with a noiseband carrier (Shannon *et al.*, 1995). In vocoder studies, one can meaningfully use 2 to 20 spectral channels before introducing other distortions, for example, from filters (Başkent, 2006; Stone *et al.*, 2008). However, if the purpose is to implement an acoustic simulation of CIs, usually the limited range of four to eight channels is preferred. The reason for this range is that four to eight channel simulations tested with normal-hearing adults produce speech perception performance most similar to that of adult CI users (Friesen *et al.*, 2001) and four to six channel simulations with normal-hearing adults to that of pediatric CI users (Dorman *et al.*, 2000; exact age of children tested not reported).

Only a small number of studies explored children's perception of speech materials that were spectrally degraded using a noiseband vocoder. The results varied depending on children's age, the speech materials used, and the degree of spectral degradation applied. Using simple, meaningful, and age-appropriate words, Newman and Chatterjee (2013) showed that toddlers between 26 and 28 months old could perceive these words successfully with 8 and 24 channel degradations. However, these toddlers needed longer processing times at 8 channels than at 24 channels. Dorman *et al.* (2000) observed that young children, from around 4 to 8 yrs old, had difficulties with understanding spectrally degraded words. They needed better spectral resolution (12 spectral channels) than adults (10 spectral channels) to reach ceiling performance. Eisenberg *et al.* (2000) observed that perception of 4- to 8-channel vocoded syllables, words, and high-context sentences was the same in older children, aged 10 to 12 yrs, as in adults, but significantly poorer in younger children, aged 5 to 7 yrs. The younger children, again, generally needed a higher spectral resolution (16 to 32 channels, depending on the speech materials) than older children and adults (4 to 8 channels) to reach ceiling performance. Eisenberg *et al.* (2000) attributed this finding to younger children not being able to make effective use of the sensory information, as well as their not yet fully developed linguistic and cognitive skills. Nittrouer *et al.* (2009) used syntactically correct but semantically anomalous short sentences, and also observed lower recognition of 4- and 8-channel degraded materials by children aged 7 yrs than by adults. However, when they used a control condition with a different manipulation of the materials, namely replacing the lowest three vocal-tract resonances with sine waves, this difference disappeared. Nittrouer *et al.* (2009) interpreted these findings as indicating that the young children did not suffer from a general immature auditory system, but instead, that children learn to recognize and make use of different speech cues at different ages. The two simulation methods they used preserve different parts of the speech signal: The noiseband vocoder preserves the amplitude envelopes, but in a reduced number of spectral channels (e.g., four or eight), whereas the sine wave speech preserves the spectral structure but reduces the rest of the speech signal cues. Children did not do as well as adults in the former case, but performed adult-like in the latter. This observation was taken as a suggestion that children are good at using spectral cues, but not yet good at using (reduced number of) amplitude cues.

The present study was designed based on previous literature on both the DPBE and the vocoder processing with children. Younger (5 to 8 yrs) and older (10 to 11 yrs) children were tested and compared to a control group of adults because within these two age ranges substantial changes were observed in both perception of reflexives and pronouns (Koster, 1993; Van Rij *et al.*, 2010) and spectrally degraded speech materials (Dorman *et al.*, 2000; Eisenberg *et al.*, 2000; Nittrouer *et al.*, 2009; Newman and Chatterjee, 2013). To keep the results relevant to actual pediatric CI users, spectral degradations of four and eight channels were used.

II. METHODS

A. Participants

Two groups of children were tested: 56 younger children (mean age 6;5 yrs, range 5;0–8;1; 33 boys) and 15 older children (mean age 10;7, range 10;1–11;6; 3 boys). In addition, a control group of 22 adults (mean age 21 yrs, range 19–26; 7 men) was tested. All participants were native monolingual speakers of Dutch with no known speech, language, or development problems. For the child participants this assessment was based on parent- and school-reports, and for the adult participants on self-reports. All were normal hearing in both ears; this was confirmed with a modified audiometric test using a portable screening audiometer AS608/AS608e (serial 744081) and using a sound insulating DD45 Audiometric Headset (Interacoustics, Assens, Denmark) in a quiet room. Based on thresholds measured from the better ear, hearing thresholds of 20 dB hearing level (HL) or better at audiometric test frequencies between 250 and 3000 Hz, with all hearing thresholds not exceeding 25 dB HL, were accepted as an indication for normal hearing. The reason for testing a reduced range of frequencies compared to standard audiometric testing was the limited testing time due to conducting the tests at schools, where children had to leave the classroom to participate in the experiment, as well as the (age-normal) limited attention span of children. We allowed thresholds up to 25 dB HL because we wanted to take into account that the tests were not conducted in special sound-insulated testing rooms. Even though the experimenter made sure the settings were quiet, there were occasional sounds in the background, which may have caused a slight shift in thresholds.

Nineteen children who were tested were eventually excluded from data analysis because they were bilingual (3), were diagnosed with ADHD (1), showed high thresholds in the audiometric screening test (12), or made more than one error on the four control items used for further screening (3). Also seven adults who were tested were eventually excluded from data analysis because they showed high thresholds in the audiometric screening test.

The children received a small gift (e.g., stickers, puzzle books) for their participation; all adults were university students who volunteered or received course credit for their participation.

All participants were tested with normal non-degraded speech, but only with one spectral-degradation condition of the CI simulation (either four or eight channels). The child and adult groups were both semi-randomly separated into two CI simulation groups, while also attempting to balance age and gender. In the children's group (i.e., the 5 to 8 yr olds and the 10 to 11 yr olds taken together), the 4-channel degradation group ($n = 38$, 21 boys; mean age = 7;3) did not differ in age from the 8-channel degradation group ($n = 33$, 15 boys; mean age = 7;5), as shown by Welch two sample test ($t(68.7) = 0.22$; $p = 0.82$).

B. Task

Children's and adults' interpretations of reflexives and pronouns were tested using a Picture Verification Task. In this task participants were presented with a (pre-recorded) test sentence and a picture illustrating a potential interpretation of that test sentence (see Fig. 1 for examples). Participants were asked to verify whether the sentence was a correct description of the picture or not by pressing one of two buttons (a button with a happy smiley face representing a yes-response and a button with a sad face representing a no-response). By presenting participants with only one picture at a time, this task tested whether the illustrated interpretation was acceptable or not, rather than whether it was the preferred interpretation for the sentence.

C. Stimuli

Speech stimuli were the original recordings used by van Rij *et al.* (2010). The sentences were spoken by a female native-Dutch speaker at normal speech rate (on average 4.1 syllables/s), and recorded monaurally at 44.1 kHz. The sentences were pronounced with normal intonation, without stressing the reflexive or the pronoun.

Two types of test sentences were recorded: Sentences with a reflexive (*himself*) and the same sentences with a pronoun (*him*):

> "Kijk, een pinguïn en een schaap zijn op de stoep. De pinguïn slaat zichzelf/hem met een pan."
> (Translation: *Look, a penguin and a sheep are on the sidewalk. The penguin is hitting himself/him with a pan.*)

Each test sentence was preceded by an introductory sentence in which the two animals shown in the picture were introduced.

The test items were 32 sentence-picture pairs, divided over two speech blocks: A block with normal speech with no degradation and another block with degraded speech. Each block included eight sentence-picture pairs with a reflexive and eight sentence-picture pairs with a pronoun. Half of these sentences were combined with a matching picture (illustrating a correct adult-like interpretation), and the other

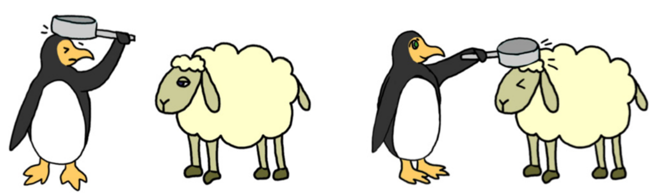

FIG. 1. (Color online) Two example pictures used in the experiment. Each picture was combined with a spoken sentence containing either a reflexive or a pronoun (see Sec. II C for a description of the speech stimuli), for example: "Kijk, een pinguïn en een schaap zijn op de stoep. De pinguïn slaat zichzelf/hem met een pan." (Translation: Look, a penguin and a sheep are on the sidewalk. The penguin is hitting himself/him with a pan.) The sentence containing a reflexive is a correct description of the picture displaying a self-oriented action (left) and is an incorrect description of the picture displaying another-oriented action (right). This pattern is the other way around if the sentence contains a pronoun rather than a reflexive.

half were combined with a non-matching picture (illustrating an incorrect non-adult-like interpretation). This was done to address the "yes" bias by children. In a verification task where children have to indicate whether the sentence they hear matches a given picture or not, children prefer to answer yes and hence perform better in match situations (where the target response is "yes") than in mismatch situations (where the target response is "no"; e.g., Chien and Wexler, 1990; Van Rij *et al.*, 2010). By including an equal number of matching and non-matching pictures, the yes bias is neutralized. A summary of the experimental items is shown in Table I.

At the beginning of the degraded speech block only, participants received two warm-up items to introduce them to vocoder processing. For this, the two sentences "Hier zie je een poes en een varken." (translation: *Here you see a cat and a pig*.) and "Hier zie je een schaap en een schildpad." (translation: *Here you see a sheep and tortoise*.) were used. The participants did not have to give a response for the two warm-up items. Furthermore, preceding the test items in each block, two practice items were provided. The practice items consisted of an introductory sentence and a simple intransitive or transitive sentence without a pronoun or reflexive, and should be easy to interpret even by the youngest children in the study. A sample practice item is "Kijk, een vogel en een paard zijn in de tuin. De vogel zit op een tak." (Translation: *Look, a bird and a horse are in the garden. The bird is sitting on a branch*.) These practice items were only used to familiarize the participants with the task and they were not included in the data analysis.

In each block, interspersed with the 16 test items, there were 4 control items. These control items were included to make sure the participant understood the task. The control items were similar to the practice items, but these were used for the sole purpose of screening, such that only data from children who correctly identified three or more items were included in final analysis (based on van Rij *et al.*, 2010). This allowed us to exclude children who still did not understand the task in spite of the practice items or who failed to pay attention during the task.

Thus, each participant received 46 items in total.

The stimuli were presented through a laptop (HP Compaq 6730b (Hewlett-Packard, Palo Alto)) with a soundcard (SoundMAX HD Audio (ADI, Norwood)) and diotically over Sennheiser HD 429 headphones (Sennheiser, Wedemark, Germany). In a pilot study, the appropriate volume setting of the laptop was determined such that all stimuli were audible and comfortable for children. This presentation level was measured to be around 60 dB sound pressure level using the KEMAR head (G.R.A.S., Holte, Denmark) and the SVANTEK 979 sound level meter (Svantek, Warsaw, Poland). The same level was used throughout the experiment. The pictures were taken from the materials used in earlier experiments (Spenader *et al.*, 2009; Van Rij *et al.*, 2010). They showed two cartoon-like animal characters that were familiar to 4-yr-old children (Fig. 1). Both animal characters were presented at approximately the same size, to avoid differences in visual saliency.

D. Cochlear-implant simulations

The CI processing was simulated offline with a four or eight channel noiseband vocoder implemented using MATLAB software, similar to the study by Başkent (2012). With a pilot study, we confirmed that the four-channel degradation was not too difficult for 5 to 8 yr-old children and they could perform above floor level. The stimuli were first bandpass filtered into a spectral range of 150 to 7000 Hz, then divided into 4 or 8 analysis bands using third order Butterworth bandpass filters with cutoff frequencies that simulate equal cochlear distances. Envelopes from these filter outputs were extracted using half-wave rectification and low-pass filtering with a third order Butterworth low-pass filter with a cutoff frequency of 160 Hz. Carrier noise bands were produced by filtering

TABLE I. Summary of the practice, control, and test items for each data collection block (normal speech and degraded speech). The practice items were presented in the same order at the beginning of each block. The order of the other items was randomized within each block. All test items consisted of an introductory sentence, a test sentence, and a picture, except for the warm-up items at the beginning of the degradation block, which consisted of an introductory sentence and a picture, but no test sentence.

Condition	Block normal non-degraded speech	Block spectrally degraded speech (four or eight channels)
Warm-up items	None	Two introductory sentences only, no reflexive/pronoun, degraded speech
Practice items	Two practice items, no reflexive/pronoun, non-degraded speech	Two practice items, no reflexive/pronoun, degraded speech
Control items	Four control items, no reflexive/pronoun, non-degraded speech	Four control items, no reflexive/pronoun, degraded speech
Matching reflexives	Four test items with reflexive, non-degraded speech, self-oriented pictures	Four test items with reflexive, degraded speech, self-oriented pictures
Non-matching reflexives	Four test items with reflexive, non-degraded speech, other-oriented pictures	Four test items with reflexive, degraded speech, other-oriented pictures
Matching pronouns	Four test items with pronoun, non-degraded speech, other-oriented pictures	Four test items with pronoun, degraded speech, other-oriented pictures
Non-matching pronouns	Four test items with pronoun, non-degraded speech, self-oriented pictures	Four test items with pronoun, degraded speech, self-oriented pictures
Total	Two practice items + 4 control items + 16 test items = 22 items	Two warm-up items + 2 practice items + 4 control items + 16 test items = 24 items

white noise with synthesis filters that were identical to analysis filters and modulating them with the corresponding envelope. The summed modulated carrier bands constituted the processed stimuli.

E. Procedure

The child participants were tested in a quiet room at their school. They could give their response by pushing one of two buttons on a keyboard. Due to the "yes" bias in children, explained before, "no" answers can be used as a double check. Here, when a child responded no, regardless of the accuracy of this answer, the child was asked to explain to a hand puppet why the sentence did not match the picture. In a small number of occasions, the child indicated to the experimenter that they had not heard the experimental sentence due to external circumstances. This happened only twice during the entire data collection with the younger group of 5 to 8 yr old children, and never with the older group of 10 to 11 yr old children. In these rare cases, the items were rerun. Further, in the response phase of each trial, children were able to revise their answer. In data analysis, only these final answers were taken into account.

The children were semi-randomly divided into two groups, balanced for age and gender: One group of children (regardless of channel condition, gender, and age) would hear the block of normal sentences first and degraded sentences second (normal + degraded), and the other group would hear the block of degraded sentences first and normal sentences second (degraded + normal). Each participant heard the same normal and degraded sentences; the normal sentences were similar but not identical to the degraded sentences, so that there would be no learning effect between blocks. Within each block, practice items were presented first, and reflexive items were interspersed with pronoun items and control items (yielding a total of 20 items), presented later. We used two different orders of items within each block (order 1 and order 2), distributing the different types of test items in a semi-random order. This resulted in a total of four versions of the experiment (A: Normal + degraded, order 1; B: Degraded + normal, order 1; C: Degraded + normal, order 2; D: Normal + degraded, order 2). On average, the hearing screening with the children took 10 to 15 min, while the comprehension task took about 10 min.

The adult group was tested in a quiet room at the university. Except for the hand puppet, the materials and procedure were exactly the same as for the children.

F. Statistical analysis

The results were reported in raw percent correct scores for visual inspection only. These were converted into d' scores to account for the yes bias of children, and statistical analyses were conducted on the d' scores using linear mixed-effects models (cf. Baayen *et al.*, 2008). The categorical predictors *Age Group* (young children, older children, and adults), *Type* (reflexive, pronoun), and *Spectral Degradation* (Four-channel, eight-channel, none), and *Block* (first part, second part of experiment) were included as fixed effects. Random intercepts were included for *Participants*. *Type*, *Degradation*, and the interaction between these two were included as random slopes per participant. This was the maximum random effects structure supported by the data.

III. RESULTS

Figure 2 shows the percentage correct responses for all three age groups (columns) and the two sentence types (rows). These raw scores are presented only to give an impression of the overall performances in percentage correct scores.

The main test used for the experiment was a "yes/no" paradigm on a one-interval discrimination task. To take into account the yes bias of children, the results were analyzed

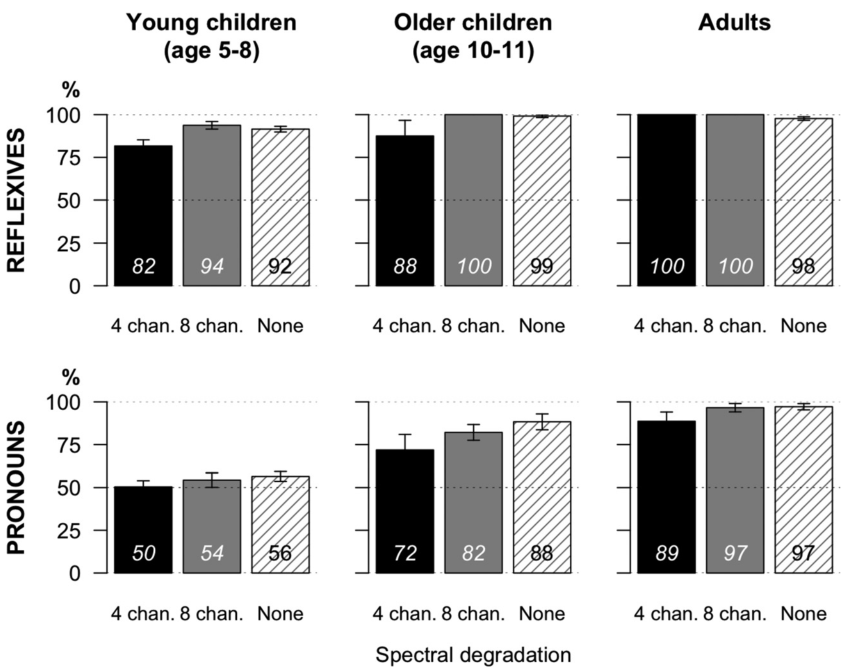

FIG. 2. Average mean percentage correct responses shown for all age groups for the comprehension of reflexives and pronouns, as a function of spectral degradation. The error bars denote one standard error of the mean.

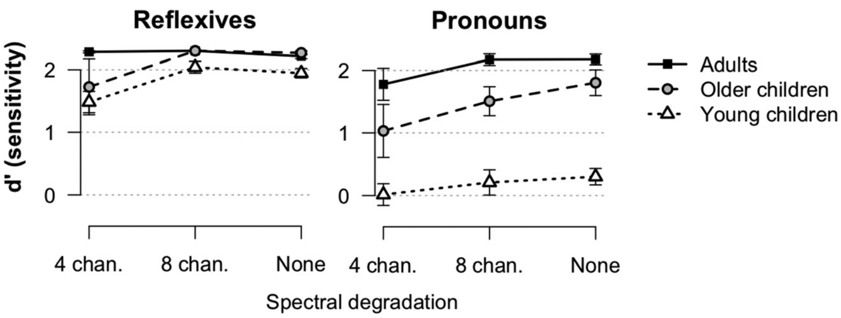

FIG. 3. Average d' scores shown for all age groups for the comprehension of reflexives and pronouns, as a function of spectral degradation. The error bars denote one standard error of the mean.

and presented in d', a measure of sensitivity for detecting a signal (based on signal detection theory; e.g., Macmillan and Creelman, 1991; Stanislaw and Todorov, 1999). Figure 3 shows the average d' scores, based on the difference between the normalized proportions of "yes" answers on the match trials (hits) and the "yes" answers on the mismatch trials (false alarms) for reflexives and pronouns (left and right panels, respectively) for all age groups. Recall that all participants were tested with normal speech, but only with one condition of vocoder processing (four or eight channels). Therefore, the normal condition in Fig. 3 ("None", i.e., no spectral degradation) was calculated by collapsing the data from all participants within each age group, while the degraded conditions reflected only the participants who were tested with the specific degradation condition.

To investigate the effect of spectral degradation on the interpretation of reflexives and pronouns for the different age groups, the d' scores [Φ^{-1}(hit rate) - Φ^{-1}(false alarm rate)] per participant per sentence type (reflexive or pronoun), and spectral degradation level (four channel or eight channel, compared with normal speech) were analyzed. A large d' value suggests a high sensitivity, because the hit rate is high, while the false alarm rate is low. A d' value close to zero means that hit rate and false alarm rate are almost equal, suggesting difficulty to distinguish between reflexive or pronoun items. Extreme values of hit and false alarm rates (0 and 1) were replaced: Rates of 0 were replaced with $0.5/n$, and rates of 1 were replaced with $(n - 0.5)/n$ (cf. Stanislaw and Todorov, 1999). Therefore, the normalized proportions ranged between -1.15 and 1.15, producing a maximum $d'_{max} = \max(\Phi^{-1}(\text{hit rate})) - \max(\Phi^{-1}(\text{false alarm rate})) = 2.30$.

First, the data are visually inspected. On the right side of each panel the performance of all groups with normal non-degraded speech is shown. With normal speech, an age effect was observed on perception of reflexives and pronouns, in a way as would be expected from the literature. With reflexives, older children performed as well as adults ($d' = 2.27$). Younger children had a lower performance than adults, yet this performance level was still high ($d' = 1.95$) and can be said to be "adult-like" (in accordance with the way children's performance on linguistic tasks is generally assessed, as was explained in Sec. I). With pronouns, younger children were close to chance level ($d' = 0.30$), showing almost no discrimination for pronouns. Older children, while lower than adults, still showed a high level performance ($d' = 1.80$, again in accordance with adult-like performance).

The effect of spectral degradation can be seen in the middle and leftmost conditions in each panel. These showed that the patterns observed with normal speech did not change drastically with eight-channel degradation. Four-channel degradation caused a small drop in d' scores for reflexives and pronouns in both young ($d' = 1.48$ and $d' = 0.02$, respectively) and older ($d' = 1.73$ and $d' = 1.03$, respectively) children, and only for pronouns in adults ($d' = 1.78$).

Second, statistical analyses were conducted on d' scores, as was explained in Sec. II. Statistics confirmed visual inspections (see the Appendix for full analysis). Using a step-wise model comparison procedure, a significant two-way interaction was found between *Age Group* and *Type* ($\chi^2(2) = 43.89$; $p < 0.001$), but the other interactions did not reach significance. A significant main effect was found for *Spectral Degradation* ($\chi^2(2) = 21.73$; $p < 0.001$). On reflexive items, the best-fitting model, including the two-way interaction *Age Group* $\times$ *Type*, suggests that the older children do not differ from adults in sensitivity, but the younger children show lower sensitivity than adults and older children ($\beta = -0.12$; standard error (SE) $= 0.049$; $t = -2.40$). On pronoun items, the difference between young children and adults is even larger ($\beta = -1.72$; SE $= 0.19$; $t = -9.23$), and also older children show lower sensitivity than adults ($\beta = -0.55$; SE $= 0.25$; $t = -2.23$). The adults did not show a significant difference in sensitivity between pronoun and reflexive items ($\beta = -0.039$; SE $= 0.16$; $t = -0.24$). In addition to the effects of *Age Group* and *Type*, all age groups show a lower sensitivity with four-channel degraded speech than with normal speech ($\beta = -0.41$; SE $= 0.091$; $t = -4.42$), but a slightly higher sensitivity on average with eight-channel degraded speech in comparison with normal speech ($\beta = 0.051$; SE $= 0.024$; $t = 2.12$).

To summarize, a correct interpretation of pronouns is acquired much later in language development (and is adult-like only at age 10 to 11) than a correct interpretation of reflexives (which is adult-like at age 5 to 8 and fully acquired at age 10 to 11). Eight-channel degradation does not affect the results, while four-channel degradation decreases the interpretation of reflexives and pronouns, but the relative patterns between age groups stays similar to that of eight channel degradation and no degradation.

IV. DISCUSSION

The aim of the present study was to explore children's comprehension of reflexives and pronouns as a potential linguistic test that can be used in young CI users aged 5 yrs and older. As a first step, we measured the effects of age and reduced spectral degradation on the comprehension of reflexives and pronouns by testing normal-hearing younger (5 to 8 yrs old) and older (10 to 11 yrs old) children, as well as a control group of adults.

A first finding was that the age effect reported in the DPBE literature was confirmed by the non-degraded normal speech condition in the present study. The DPBE literature suggests that children reach adult-like comprehension (i.e., performance higher than ca. 90% correct) of reflexives around the age of 5 yrs, and adult-like comprehension of pronouns around the age of 10 yrs (Koster, 1993; van Rij *et al.*, 2010). In our study, we had selected younger and older children around these ages, namely 5 to 8 yrs and 10 to 11 yrs, respectively, to be able to investigate these linguistic markers. Indeed, the younger children showed adult-like performance for reflexives only, while the older children showed adult-like performance for both reflexives and pronouns, as was expected from the DPBE literature. Further, younger children were close to chance level with pronouns, indicating that pronoun comprehension gradually develops between the ages of 5 and 10 yrs. The DPBE, then, is confirmed to give rise to two related linguistic markers at different ages that can potentially be used in young CI users to test language development.

A second finding was the effect of spectral degradation on children's comprehension of reflexives and pronouns. Eight-channel degradation did not affect interpretation much compared to normal speech, neither in the young and older groups of children, nor in the adult control group. It was only with four-channel degradation that the performance started to decrease, and this effect was observed, again, with both groups of children.

Finally, our study showed that degraded speech had a different effect on the interpretation of the two types of stimuli. For the adults, pronoun items were affected by degraded speech, but reflexive items were not. Apparently, reflexive items are highly robust under acoustically challenging conditions, at least for the stimuli used in our study. This might be because we used the reflexive *zichzelf*, which is longer and more complex than the pronoun *hem* and also tends to be stressed. As a consequence, it is perceptually more salient. In a future study, it would be interesting to see whether the same robustness would be found if, instead of the longer reflexive *zichzelf*, the shorter and unstressed reflexive *zich* was used. Because we took care that the pronoun was not stressed in the pre-recorded sentences, as would be most natural in the context of those sentences, the pronoun may have been extra susceptible to the effects of our speech degradation manipulation. But note that with the linguistic materials employed in the present study, in particular the transitive verbs in the test sentences such as *hit*, the unstressed reflexive *zich* could not have been used, as it would not have been grammatically correct. The more often a verb is used to refer to a self-oriented action, the more often the unstressed reflexive *zich* is used with this verb (e.g., Smits *et al.*, 2007). As *wash* and *shave* more often refer to a self-oriented action than *hit* and *tickle*, the former verbs allow the unstressed reflexive more easily. However, including the former type of verbs would also imply introducing a bias that the sentence most likely refers to a self-oriented action, which could potentially cause children to make even more errors with the pronoun *hem* than they already did in the present study.

How do our findings compare to previous studies? The study by Eisenberg *et al.* (2000) employed children of similar ages (young children aged 5 to 7 yrs, and older children aged 10 to 12 yrs), who were tested with speech materials spectrally degraded with vocoder processing with 4 channels and up. The test materials used were syllables, words, and meaningful and high-context sentences. Eisenberg *et al.* observed that in non-degraded conditions all children performed well, at a level comparable to that of adults. However, with spectral degradations, only the younger children performed worse than adults. The older children dealt with degradations well, performing almost at the level of adults. These results were interpreted as indicating that speech perception mechanisms are still developing in the younger children.

Our study targeted language developmental milestones in children of similar ages. Our younger children seemed to not have acquired the correct interpretation of pronouns yet, but they had done so with reflexives. Our older children, in contrast, seemed to have mastered the correct interpretation of both linguistic forms. With age, it appears that listeners become less vulnerable to degradation of speech with linguistic forms that they have already acquired. Perhaps this is because they have also developed the linguistic skill to *predict* these linguistic forms in a sentence context in a top-down manner and hence are less dependent on bottom-up information in the speech stream to actually perceive these forms. As for the specific effects of spectral degradation, the pattern of the effect of spectral degradation in the present study was different than what Eisenberg *et al.* observed. In the present study, the effect of spectral degradation of both four and eight channels as compared to normal speech was similar in the two groups of children. Hence, we did not observe a special deficit with younger children in dealing with spectrally degraded linguistic stimuli. Instead, on both groups and in a similar way, there was no effect of eight-channel degradation and a mild effect of four-channel degradation.

This difference in findings between our study and the study of Eisenberg *et al.* could be due to the nature of the speech materials used. In the present study, we used stimuli that are specifically geared toward identifying the language developmental milestones for children's comprehension of reflexives and pronouns. Perhaps once children reach these milestones, roughly at the ages of 5 and 10 yrs, respectively, degradation in signal quality produces a simple linear effect on perception performance, perhaps due to a reduction of overall speech cues in the impoverished signal. Nittrouer *et al.* (2009), supporting this idea, claimed that the reduction in the perception of noiseband vocoded speech they observed appeared to be a consequence of the cues that were preserved from the original signal. These reduced cues would be a number of slowly varying amplitude envelopes from a limited number of spectral channels, while the fine temporal and spectral information is removed, and Nittrouer *et al.* (2009) had suggested that the use of these cues by children is perhaps not completely mastered. This may also be the case in our situation, as spectral degradation caused a slight shift in the performance of younger and older children, without changing the overall linguistic milestone patterns drastically. With normal speech without spectral degradation, younger as well as older children performed well with the interpretation of reflexives, younger children could not discriminate pronouns yet, and older children performed well with the interpretation of pronouns. This pattern did not change with the application of spectral degradation, which only caused a slight reduction in the performance, but in a similar way for both younger and older children.

The findings of the present study have consequences for a potential application of the test in characterizing language development in CI children. These children could have delayed language development depending on the age of implantation (e.g., Kirk *et al.*, 2002), while, at the same time, they have to deal with impoverished signal delivery from the CI, as was simulated by the spectral degradation of this study. We propose that the linguistic test may best characterize the slower language development and ideally only shows minimal effects of impoverished signal delivery, and the comprehension of reflexives and pronouns seems to be a good test in this sense. Assuming that four- and eight-channel simulation indeed captures a realistic range of degradations that can occur in actual CIs (e.g., Dorman *et al.*, 2000; Friesen *et al.*, 2001), our study confirms that the signal degradation imposed by CI speech transmission has only a slight effect on the results. Consequently, for all children older than 5 yrs of age, non-adult-like comprehension of reflexives would point at a delay in language development. Similarly, for all children older than 10 yrs of age, non-adult-like comprehension of pronouns would point at a delay in language development. Hence, the test potentially provides a behavioral method rather than a subjective questionnaire, which a clinician can quickly use to identify a potential delay in these linguistic milestones.

In summary, observing the differences in results between previous studies with simple speech materials and the present study testing children's and adults' comprehension of reflexives and pronouns in various speech degradation conditions, we propose that the comprehension of reflexives and pronouns can provide additional information on linguistic milestones for CI children that the standard clinical speech perception tests may not fully reveal.

ACKNOWLEDGMENTS

This investigation was supported by VIDI and VICI grants from the Netherlands Organization for Scientific Research, NWO, awarded to Deniz Başkent (Grant No. 016.096.397; from Netherlands Organization for Health Research and Development, ZonMw) and Petra Hendriks (Grant No. 277-70-005), respectively. Further support came from a Rosalind Franklin Fellowship from the University of Groningen, University Medical Center Groningen, and funds from the Heinsius Houbolt Foundation. We thank OJBS De Vuurvogel, the Maria in Campisschool and day care center Pinokkio in Assen and the Van Heemskerckschool in Hoogezand, all children and their parents/caretakers, teachers, group leaders, and adult participants for their participation in the experiment. We are also grateful to Nick Becker, Aurora Sayers, and Inge Wessel for their help in collecting the data, and Bert Maat and Anne Keegstra for their general support. The study is part of the research program of the Otorhinolaryngology Department of University Medical Center Groningen: Healthy Aging and Communication.

APPENDIX: SUMMARY OF BEST-FITTING STATISTICAL MODEL

Model: dp ~ AgeGroup + Type + Speech + AgeGroup: Type + Block + (1 | Subject) + (0 + Type * Speech | Subject)

Random effects		Variance	Std. Dev.
Subject	(Intercept)	1.708×10^{-5}	0.004
Subject	$\text{Type}_{\text{PRONOUN}}$	6.303×10^{-1}	0.794
	$\text{Type}_{\text{REFLEXIVE}}$	1.238×10^{-1}	0.352
	$\text{Speech}_{4\text{CHAN}}$	1.231×10^{-1}	0.351
	$\text{Speech}_{8\text{CHAN}}$	1.368×10^{-2}	0.117
	$\text{Type}_{\text{PRONOUN}}:\text{Speech}_{4\text{CHAN}}$	1.175×10^{-1}	0.343
	$\text{Type}_{\text{PRONOUN}}:\text{Speech}_{8\text{CHAN}}$	5.164×10^{-2}	0.227

Fixed effects	Estimate (β)	Std. Error	*t* value
(Intercept)	2.160	0.057	38.21
$\text{AgeGroup}_{\text{CHILDREN}}$	−0.117	0.049	−2.40
$\text{AgeGroup}_{\text{OLDERCHILDREN}}$	0.000	0.064	0.00
$\text{Type}_{\text{PRONOUN}}$	−0.039	0.159	−0.24
$\text{Speech}_{4\text{CHAN}}$	−0.406	0.092	−4.42
$\text{Speech}_{8\text{CHAN}}$	0.052	0.024	2.12
Block	0.032	0.023	1.37
$\text{AgeGroup}_{\text{CHILDREN}}:\text{Type}_{\text{PRONOUN}}$	−1.723	0.187	−9.23
$\text{AgeGroup}_{\text{OLDERCHILDREN}}:\text{Type}_{\text{PRONOUN}}$	−0.553	0.248	−2.23

Procedure: After determining the maximum random effects structure allowed by the data, a stepwise model comparison procedure was used for determining the best-fitting model, starting with the full interaction model. The contrasts of all predictors were centered. The table above presents the best-fitting model with better interpretable contrasts:

$\text{AgeGroup}_{\text{CHILDREN}}$: Comparison between the young children (1) and adults (0).

$\text{AgeGroup}_{\text{OLDERCHILDREN}}$: Comparison between older children (1) and adults (0).

$\text{Type}_{\text{PRONOUN}}$: Comparison between pronouns (1) and reflexive (0) items.

$\text{Speech}_{4\text{CHAN}}$: Comparison between four-channel degraded speech (1) and normal speech (0).

$\text{Speech}_{8\text{CHAN}}$: Comparison between eight-channel degraded speech (1) and normal speech (0).

Baayen, R. H., Davidson, D. J., Bates, D. M. (**2008**). "Mixed-effects modeling with crossed random effects for subjects and items," J. Mem. Lang. **59**, 390–412.

Başkent, D. (**2006**). "Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels," J. Acoust. Soc. Am. **120**, 2908–2925.

Başkent, D. (**2012**). "Effect of speech degradation on top-down repair: Phonemic restoration with simulations of cochlear implants and combined electric-acoustic stimulation," J. Assoc. Res. Otolaryngol. **13**, 683–692.

Chien, Y. C., and Wexler, K. (**1990**). "Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics," Lang. Acquisition **1**, 225–295.

Dorman, M. F., Loizou, P. C., Kemp, L. L., Kirk, K. I. (**2000**). "Word recognition by children listening to speech processed into a small number of channels: Data from normal-hearing children and children with cochlear implants." Ear. Hear. **21**, 590–596.

Eisenberg, L. S., Shannon, R. V., Martinez, A. S., Wygonski, J., and Boothroyd, A. (**2000**). "Speech recognition with reduced spectral cues as a function of age," J. Acoust. Soc. Am. **107**, 2704–2710.

Friesen, L. M., Shannon, R. V., Başkent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Geers, A. E., and Nicholas, J. G., (**2013**). "Enduring advantages of early cochlear implantation for spoken language development," J. Speech Lang. Hear. Res. **56**, 643–653.

Kirk, K. I., Miyamoto, R. T., Lento, C. L., Ying, E., O'Neill, T., and Fears, B. (**2002**). "Effects of age at implantation in young children," Ann. Otol. Rhinol. Larynol. Suppl. **111**, 69–73.

Koster, C. (**1993**). "Errors in anaphora acquisition," OTS dissertation series, Onderzoeksinstituut voor Taal en Spraak, Rijksuniversiteit Utrecht, The Netherlands, 191 p.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge), 407 p.

McKee, C. (**1992**). "A comparison of pronouns and anaphors in Italian and English acquisition," Lang. Acquisition **2**, 21–54.

Newman, R., and Chatterjee, M. (**2013**). "Toddlers' recognition of noise-vocoded speech," J. Acoust. Soc. Am. **133**, 483–494.

Nicholas, J., and Geers, A. (**2006**). "Effects of early auditory experience on the spoken language of deaf children at 3 years of age," Ear Hear. **27**, 286–298.

Nilsson, J. J., Soli, D. D., and Gelnett, D. J. (**1996**). *Development of the Hearing in Noise Test for Children (HINT-C)* (House Ear Institute, Los Angeles, CA).

Nittrouer, S., Lowenstein, J. H., and Packer, R. R. (**2009**). "Children discover the spectral skeletons in their native language before the amplitude envelopes," J. Exp. Psychol. Hum. Percept. Perform. **35**, 1245–1253.

Philip, W., and Coopmans, P. (**1996**). "The double Dutch delay of Principle B effect," in *Annual Boston University Conference on Language Development. The Proceedings of the 20th Annual Boston DPBE and Language Acquisition in Down Syndrome University Conference on Language Development*, Vol. 2 (Cascadilla Press, Somerville, MA), pp. 576–587.

Reynell, J., and Gruber, C. (**1990**). *Reynell Developmental Language Scales: U.S. Edition* (Western Psychological Services, Los Angeles, CA), 67 p.

Rozendaal, M. I. (**2008**). "The acquisition of reference: A cross-linguistic study," Vol. 197 of LOT international series, Doctoral thesis, University of Amsterdam, 415 p.

Ruigendijk, E., Friedmann, N., Novogrodsky, R., and Balaban, N. (**2010**). "Symmetry in comprehension and production of pronouns: A comparison of German and Hebrew," Lingua **120**, 1991–2005.

Schorr, E. A., Roth, F. P., and Fox, N. A. (**2009**). "Quality of life for children with cochlear implants: Perceived benefits and problems and the perception of single words and emotional sounds," J. Speech. Lang. Hear. Res. **52**, 141–152.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Sharma, A., Dorman, M. F., and Spahr, A. J. (**2002**). "A sensitive period for the development of the central auditory system in children with cochlear implants: Implications for age of implantation," Ear Hear. **23**, 532–539.

Smits, E. J., Hendriks, P., and Spenader, J. (**2007**). "Using very large parsed corpora and judgment data to classify verb reflexivity," in *Anaphora: Analysis, Algorithms and Applications, 6th Discourse Anaphora and Anaphor Resolution Colloquium, DAARC2007, LNAI (Lecture Notes in Artificial Intelligence)*, edited by António Branco, #4410 (Springer-Verlag, Berlin Heidelberg), pp. 77–93.

Spenader, J., Smits, E. J., and Hendriks, P. (**2009**). "Coherent discourse solves the pronoun interpretation problem," J. Child Lang. **36**, 23–52.

Stanislaw, H., and Todorov, N. (**1999**). "Calculation of signal detection theory measures," Behav. Res. Methods Instrum. Comput. **31**, 137–149.

Stone, M. A., Füllgrabe, C., and Moore, B. C. J. (**2008**). "Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region," J. Acoust. Soc. Am. **124**, 2272–2282.

Svirsky, M. A., Teoh, S. W., and Neuburger, H. (**2004**). "Development of language and speech perception in congenitally, profoundly deaf children as a function of age at cochlear implantation," Audiol. Neuro-Otol. **9**, 224–233.

Van Rij J., Van Rijn H., Hendriks P. (**2010**). "Cognitive architectures and language acquisition: A case study in pronoun comprehension," J. Child Lang. **37**, 731–766.

Vermeulen, A., Hoekstra, C., Brokx, J., and Van den Broek, P. (**1999**). "Oral language acquisition in children assessed with the Reynell Developmental Language Scales," Int. J. Pediatr. Otorhinolaryngol. **47**, 153–155.