

Recognition of Spanish consonants in 8-talker babble by children with cochlear implants, and by children and adults with normal hearing

Ignacio Moreno-Torres, and Sonia Madrid-Cánovas

Citation: *The Journal of the Acoustical Society of America* **144**, 69 (2018); doi: 10.1121/1.5044416

View online: <https://doi.org/10.1121/1.5044416>

View Table of Contents: <http://asa.scitation.org/toc/jas/144/1>

Published by the *Acoustical Society of America*

Articles you may be interested in

[Effects of spectral resolution on spectral contrast effects in cochlear-implant users](#)

The Journal of the Acoustical Society of America **143**, EL468 (2018); 10.1121/1.5042082

[A laboratory investigation into the effect of quiet-time interval between aircraft noise events on overall noisiness](#)

The Journal of the Acoustical Society of America **144**, 11 (2018); 10.1121/1.5044403

[Effects of consonantal context on the learnability of vowel categories from infant-directed speech](#)

The Journal of the Acoustical Society of America **144**, EL20 (2018); 10.1121/1.5045192

[Perceptual evaluation of measures of spectral variance](#)

The Journal of the Acoustical Society of America **143**, 3300 (2018); 10.1121/1.5040484

[Temporal factors in cochlea-scaled entropy and intensity-based intelligibility predictions](#)

The Journal of the Acoustical Society of America **143**, EL443 (2018); 10.1121/1.5041468

[Development of speech rhythm in first language: The role of syllable intensity variability](#)

The Journal of the Acoustical Society of America **143**, EL463 (2018); 10.1121/1.5042083

Recognition of Spanish consonants in 8-talker babble by children with cochlear implants, and by children and adults with normal hearing

Ignacio Moreno-Torres^{a)}

Departamento de Filología Española, Campus de Teatinos s/n, Universidad de Málaga, 29071 Málaga, Spain

Sonia Madrid-Cánovas

Departamento Lengua Española y Lingüística General, Universidad de Murcia Campus de La Merced, 30001 Murcia, Spain

(Received 18 November 2017; revised 29 May 2018; accepted 8 June 2018; published online 5 July 2018)

This paper presents the results of closed-set recognition task for 80 Spanish consonant-vowel sounds in 8-talker babble. Three groups of subjects participated in the study: a group of children using cochlear implants (CIs; age range: 7–13), an age-matched group of children with normal hearing (NH), and a group of adults with NH. The speech-to-noise ratios at which the participants recognized 33% of the target consonants were +7.8 dB, –3 dB, and –6 dB, respectively. In order to clarify the qualitative differences between the groups, groups were matched for the percentage of recognized syllables. As compared with the two groups with NH, the children with CIs: (1) produced few “I do not know” responses; (2) frequently selected the voiced stops (i.e., /b, d, g/) and the most energetic consonants (i.e., /l, r, j, s, ʃ/); (3) showed no vowel context effects; and (4) had a robust voicing bias. As compared with the adults with NH, both groups of children showed a fronting bias in place of articulation errors. The factors underlying these error patterns are discussed. © 2018 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.1121/1.5044416>

[SHF]

Pages: 69–80

I. INTRODUCTION

Cochlear implants (CIs) are electronic devices that provide access to the speech signal to persons with profound auditory loss. CIs are generally very effective, particularly in the case of children who receive one or two implants before or close to their second birthday (e.g., Bouton *et al.*, 2012). Unfortunately, even in the case of early-implanted children the effectiveness of these devices is limited in noisy backgrounds (e.g., Caldwell and Nittrouer, 2013). This clear contrast between the very good outcomes in quiet and the poor outcomes in noise motivated our interest in early-implanted children.

While many studies have confirmed the quantitative impact of noise for adults and children with CIs, our understanding of how they process speech in noise is limited, especially in the case of children (for data about adults see Munson and Nelson, 2005; Chun *et al.*, 2015). For instance, it is not known whether adults and children with CIs produce the same consonant recognition errors as subjects with normal hearing (NH), or different ones (see Miller and Nicely, 1955; Moreno-Torres *et al.*, 2017). However, there are two indications that the errors produced in noise by children with CIs might be atypical. One is the evidence that, due to the technical limitations of their devices, CI users miss acoustic cues that are most helpful for subjects with NH to recognize

speech in noise (i.e., dynamic spectral cues; see Hedrick *et al.*, 2011). Another is the evidence that early phonological development is atypical in children with CIs (e.g., Moreno-Torres and Moruno-López, 2014), which raises the possibility that they differ from children with NH in how phonological knowledge guides speech recognition. Clarifying the types of errors produced by children with CIs might be relevant both to improve CI technology and for rehabilitation purposes.

The main aim of this study is determine to what point children with CIs are similar/different from children with NH in how they process speech in noise. A secondary aim is to compare children and adults with NH. To this end, we explored the errors produced in a nonsense syllable recognition task by three groups of participants, namely children with CIs, children with NH, and adults with NH. We expected that, due to the summed effect of the technical limitations of their devices and their atypical phonological skills, the errors of the children with CIs would be notably different from the errors in the children with NH, while the latter would be relatively similar to adults with NH.

A. Speech perception in noise in adults with normal hearing

From an acoustic perspective, the effect of noise consists of adding spectral-temporal information to a target signal. From a perceptual perspective, the precise effect of noise depends on whether the two signals are segregated

^{a)}Electronic mail: imoreno@uma.es

(i.e., whether the listener identifies which acoustic cues belong to the target). If the two signals are segregated, the effect of noise consists of deleting some or all of the acoustic cues, for which successful recognition will depend on the capacity of the listener to interpret the available cues (i.e., energetic masking; Hawkins, 2010). If the two signals are not segregated, the listener is presumably confronted with a heterogeneous and confounding set of acoustic information (informational masking, Brungart, 2001). The results of previous studies suggest that the most common situation in listeners with NH is that of energetic masking, and that errors are relatively systematic. For instance, when presented with a voiced stop (i.e., /b, d, g/) in noise, Spanish listeners produce the same errors as when the sonority bar signaling voice is removed experimentally (Gurlekian *et al.*, 1987; Feijóo *et al.*, 1998; Moreno-Torres *et al.*, 2017): they tend to recognize the voiceless counterpart (i.e., /p, t, k/). The opposite situation, which might involve the listener perceiving an in-existent sonority bar, is clearly less frequent. Similarly, when weak fricatives are presented in noise, listeners tend to rely on alternative cues, and the specific errors may depend on the accompanying vowel (e.g., Woods *et al.*, 2010; Moreno-Torres *et al.*, 2017). Given that the most common errors in Spanish involve those two errors (i.e., stop voicing and obstruent place of articulation), it seems at least for listeners with NH, energetic masking might be more common than informative masking.

It is also relevant for this study that not all phonetic cues are equally vulnerable to noise (e.g., Miller and Nicely, 1955; Woods *et al.*, 2010). We will consider three groups of acoustic cues: temporal, dynamic spectral (e.g., formant transitions) and static spectral (e.g., fricative noise). It is generally agreed that temporal cues are more resistant to signal degradation than spectral ones (Miller and Nicely, 1955). For spectral static cues, a distinction must be made between weak cues (e.g., the fricative noise of low energy fricatives such as /f, x.../), and robust ones (e.g., the frication of sibilants and the formants of nasals). It is important that typically, when weak spectral cues are masked by noise, listeners tend to rely on dynamic spectral cues. This is true both for stop voicing (Gurlekian *et al.*, 1987; Feijóo *et al.*, 1998; Moreno-Torres *et al.*, 2017) and for the place of articulation of fricatives (e.g., Woods *et al.*, 2010). Altogether, the above results suggest the following resistance ranking for listeners with NH: temporal > high-energy static spectral \approx dynamic spectral > low-energy-static spectral.

B. Speech perception in children

Research on speech perception development suggests that the age at which children attain adult-like levels depends on the specific task. For instance, temporal and spectral discrimination reaches adult-like performance between 4 and 5 years of age (Allen and Wightman, 1992); perception of nonsense syllables in quiet is mature around age 7 years (Hnath-Chisolm *et al.*, 1998); use of dynamic spectral cues is mature after 7 years of age (Nittrouer, 2004); categorization of manipulated signals (i.e., with some acoustic cues missing) is mature between 10 and 12 years (Gerrits, 2001; Hazan and

Barret, 2000); and phoneme identification in noise about 14 years of age (Johnson, 2000). Thus, tasks requiring the ability to process incomplete speech signals (as in the case of speech recognition in noise) seem to be the most difficult ones for children.

Different proposals have been made to explain these developmental differences. Some authors have proposed that children differ from adults in the strategies used to weight various acoustic cues (Nittrouer and Crowther, 1998; Nittrouer, 2004). Specifically, Nittrouer and colleagues propose that young children 4–7-years-old weight formant transitions more strongly than adults. This proposal is relevant for this study because formant transitions are especially relevant to recognition of speech in noise (e.g., Woods *et al.*, 2010) and are poorly encoded by CIs (Hedrick *et al.*, 2011). Other authors have claimed that, due to having limited auditory acuity, children rely heavily on cues that are louder, longer, or spectrally more informative (i.e., highly audible; Sussman, 2001; Mayo and Turk, 2004). Finally, it has also been proposed that part of the differences between adults are due to children making limited use of predictability of acoustic cues (i.e., insufficient language experience; Mayo *et al.*, 2016). Note that the different studies have provided convincing evidence of the role of each of these three factors, which suggests that children might differ from adults in all three of them: the perceptual strategies, the auditory acuity, and the ability to make acoustic predictions.

C. Speech perception with a CI

Despite the known limitations of CI technology, some studies have found clear similarities in how children with CIs and children with NH process speech (Giezen *et al.*, 2010; Nittrouer *et al.*, 2014). For instance, Giezen *et al.* (2010) analyzed four vowel contrasts and two consonant contrasts in a group of children with CIs (i.e., b/p and f/s). The authors found that the CI group used the cues in the f/s contrast less effectively than the controls, but the groups did not differ in the cue weighting strategy. Nittrouer *et al.* (2014) analyzed two contrasts (kob/kop and sa/sha). The authors did not find differences in weighting of cues to the kop–kob decision. As for the sa–sha pair, they found differences that were related to phonemic awareness and word recognition. They concluded that as regards cue recognition and weighting, the two groups were not different. In contrast with these two studies, in a study using synthetic stimuli, Hedrick *et al.* (2011) concluded that CI users struggled to recognize dynamic spectral cues. Also, Bouton *et al.* (2012) found that while children with CIs scored very close to typical for the voicing and the manner of articulation features, their scores were significantly low for the place of articulation feature. Altogether these studies indicate that there are some potentially relevant differences between subjects with CIs or with NH, though these differences might be observable only under demanding conditions (e.g., synthesized stimuli) or when specific speech cues or phonological features are involved (e.g., dynamic acoustic cues).

In contrast with the results in quiet conditions, hearing in noise with CIs is very poor (e.g., Caldwell and Nittrouer, 2013; Qazi *et al.*, 2013). Qazi *et al.* (2013) proposed that the speech coding strategies used by today's CIs may result frequently in an incoherent speech stream. That the speech stream might be incoherent raises the possibility that CI users are confronted frequently with informational masking (i.e., when unrelated acoustic cues cannot be segregated). To our knowledge, only one study has analyzed the phonemic errors of CI users in noise (Chun *et al.*, 2015). In this study, a group of Korean adult CI users and adults with NH were required to recognize a set of Korean mono-syllabic words. The participants were evaluated in quiet and at three speech-to-noise ratio (SNR) levels (i.e., 6 dB, 0 dB, and -6 dB). For the CI group, the errors in quiet were approximately 50%, and they increased up to 100% at an SNR of -6 dB. For the NH group, the errors increased from 2% to 40% for the same conditions. The errors were classified as substitution (e.g., ba > pa), omission (e.g., ba > a), addition, fail (e.g., ba > pu), and no response. For the NH group, the majority of the errors were substitutions. For the CI group, errors were more severe (e.g., fail and "no response"). Unfortunately, as this study provided few details of specific phonological errors, it is difficult to interpret them in terms of how CI users process speech.

Finally, two studies have analyzed directly phonetic processing in noise in adults with CIs. Munson and Nelson (2005) compared the performance of CI and NH listeners on the discrimination of a pair of vowels (i.e., /i -u/) and a pair of glides (i.e., /w/ - /j/) in quiet and in noise. The CI listeners performed similarly to the NH subjects for the two vowels in quiet and in noise. In contrast, on /w/ - /j/ discrimination, the CI users performed similarly to the NH controls in quiet, but significantly worse in noise. van Zyl and Hanekom (2013) compared the ability of CI recipients to discriminate question/statement intonation in the presence of speech-weighted noise to their ability to recognize vowels in the same test paradigm and listening condition. The authors found that vowel recognition was significantly better than prosody recognition in the two listener groups in both quiet and noise, and that question/statement discrimination was the most difficult task for CI listeners in noise. Given that glides such as /w/ - /j/ differ from vowels in having dynamic spectral cues, and given that statement/questions differ in the presence of ascending intonation (i.e., dynamic spectral changes), these results seem to reinforce the proposal that CI users struggle to process dynamic spectral cues. However, a more detailed description of the precise effects of noise for CI hearing is needed.

Specifically, these issues require further attention. In the first place, it seems relevant to determine whether, as in the case of NH listeners, the main effect of noise in children with CIs is that of energetic masking (i.e., associated with highly predictable errors) or, alternatively, if there is an increase in informational masking. In the second place, and assuming that the difficulties to process speech in noise will be variable for the children with CIs (i.e., some acoustic cues and sounds will be easier to process than others), it is relevant to determine the factors that may explain these

difficulties. Given the important role of dynamic spectral cues for speech processing in general and particularly for children, it seems relevant to clarify whether children with CIs use these cues or not. While previous evidence suggests that CI users fail to use formant transitions (i.e., one important type of dynamic cue) it is possible that they have access to alternative acoustic ones (e.g., spectral tilt; Alexander and Kluender, 2008). Another aspect that remains unclear is up to what point the poor auditory acuity in children with CIs may increase the relative importance of the audibility of different acoustic cues (Sussman, 2001). In a previous study with some of the participants in this study, we observed that during the first two years of implant use, the children with CIs learnt speech sounds with robust acoustic cues (e.g., sibilants) quicker than children with NH, while other speech sounds were learnt more slowly (see Moreno-Torres and Moruno-López, 2014); it remains to be determined if the effect is observed 6-8 years later in the same children. One more factor that might influence these children is the visibility of some consonants (e.g., labials are visible and hence possibly easier to learn than velars). Visual information is processed by NH listeners (McGurk and MacDonald, 1976); however, it is possible that the relative importance is increased in CI users as a compensatory strategy. Finally, it is relevant to inquire up to what point language experience influences speech processing, which would provide an indirect measure of the potential benefits of language instruction.

D. This study

This study analyzed the errors produced in a speech-in-noise task by a group of children with CIs and two groups of NH listeners (adults and children). The task consisted of recognizing nonsense consonant-vowel syllables in a background babble noise. The noise was produced by combining eight talker voices, which was expected to increase informational masking (Simpson and Cooke, 2005). After listening to each syllable, the participants could either select the option "I do not know" or one of the 16 available consonants. By allowing the participants to indicate that they had not recognized any syllable, we expected to avoid random responses which might make it difficult to interpret the results. Also, we assumed that the groups would differ in the tendency to select this option due to differences in informational masking.

Note that this study was focused on qualitative differences rather than on quantitative ones. For this reason, the data were obtained in two steps. In the first step, three groups of participants were evaluated in a task consisting of recognizing a set of 80 syllables (16 consonants \times 5 vowels) at different SNR levels (see details in the Sec. II). Next, we selected 80-syllable sets for which the mean percent correct recognition was approximately 35%. We chose this precise value for practical reasons: there were sufficient 80-syllable sets close to this value in the three groups, and we assumed that the remaining 65% errors would provide sufficient details to answer the research questions.

Using this procedure, we obtained three datasets, one per participant group (i.e., children with CIs, children with NH, and adults with NH; see details of the data selection

criteria in Sec. II). For each dataset, we computed various measures that served to answer these three questions: (1) do children with CIs, children with NH, and adults with NH produce the same phoneme error patterns? (2) Do the three groups process dynamic spectral information identically? And (3) do they show the same error biases for the phonological features of voicing and manner of articulation?

1. Phoneme error patterns

Following a long tradition in speech-in-noise studies, we used confusion matrices (CMs) to analyze phoneme error patterns (Miller and Nicely, 1955; Wang and Bilger, 1973, Dubno and Levitt, 1981; Sroka and Braida, 2005). The CMs were used to determine the different error types for the full list of consonants and for different consonant groups. The errors were classified as either omissions (i.e., “I do not know” responses) or substitutions (i.e., selecting a consonant that does not correspond to the target). We expected that due to difficulties segregating the target from the background (Qazi, 2013), the children with CIs might be less confident than the controls to discard ambiguous stimuli, which might result in a relative reduction in the ratio of “I do not know responses.”

The confusion matrices were also used to explore the role of audibility and consonant frequency. The impact of audibility was explored by comparing the number of responses (correct responses as well as false positives) for consonants classified in a previous study according to their audibility (see Moreno-Torres *et al.*, 2017): high (i.e., /j s tʃ/), middle (i.e., /m, n, l, r/), or low (i.e., /p b t d k g f θ x/). Based on previous evidence, we expected that when the mean hit rate was the same, the children with CIs would recognize more consonants with high audibility than the children with NH, and the children with NH would recognize more than the adults with NH. The existence of a frequency effect was explored using a pair of consonants whose resistance seems to be associated with frequency (i.e., /f/-/θ/). Spanish listeners typically show a bias for /θ/ (more frequent than /f/ in Spanish), while English-speaking listeners show a bias for /f/ (more frequent than /θ/). Based on previous evidence that language experience has a positive impact on speech perception tasks, we anticipated that the three groups would more easily recognize /θ/ than /f/, though the effect might be larger in the adults with NH.

2. Processing of dynamic spectral cues

We computed the percentage of correctly recognized consonants as a function of the accompanying vowel and the percentage of place of articulation errors. We assumed that a vowel effect would be observable only if the listeners took into account the transitions from the consonants to the vowels (i.e., the dynamic spectral cues) to process the consonants. Similarly, we assumed that place of articulation errors would be frequent in the CI participants if they were not able to process dynamic cues. We expected the CI users to show no vowel effect and to produce more place of articulation errors than the controls. As for the children with NH, we anticipated two alternative scenarios in comparison with the adults: (1) as children with NH pay increased attention to

dynamic spectral cues, vowel effects might be more pronounced in the children with NH than in the adults with NH (see Nittrouer, 2004); however, (2) as children with NH have less language experience than adults, the results might be more variable, which might blur any vowel effects in the children (see Mayo *et al.*, 2016).

3. Errors biases for the voicing, place and manner of articulation features

We examined the error biases for three phonological features, voicing, manner, and place of articulation, respectively. As regards to voicing, it is relevant that adults with NH show a strong bias towards devoicing of Spanish stops (i.e., $b > p / d > t / g > k$; see Moreno-Torres *et al.*, 2017). As noted above, this effect is suggestive of energetic masking of the sonority bar. As regards the CI users, we may consider two possibilities. If the phonological representations of CI users are typical, they might show the same bias as NH subjects (i.e., devoicing). Alternatively, if more audible consonants (e.g., the voiced ones; Albalá and Marrero, 1995) are better represented, the CI users might be biased towards voicing. Note also that a voicing bias might occur if the CI users wrongly interpret the background noise as a sonority bar (i.e., informational masking). We anticipated that the CI users would show a voicing bias, which might be due to the increased role of audibility for CI users and/or to informative masking. As for the children with NH, we did not expect their responses to be different from those of the adults.

As regards the manner errors, Moreno-Torres *et al.* (2017) found a relatively small percentage of errors for this feature, and a slight bias towards stopping in the labials (p/f) and dentals (t/θ), but not for the velars (k/x). The low percentage of manner errors was associated with the high resistance of duration cues. The stopping bias in the labial and dental consonants might indicate a preference for unmarked consonants and/or that background noise energetically masks more effectively the turbulences of fricatives than the stop bursts. As the manner of articulation is acquired early in both children with NH and children with CIs (Bosch, 2004; Bouton *et al.*, 2012), we expected no differences among the three groups for this feature.

For place of articulation, Moreno-Torres *et al.* (2017) did not find a fronting or backing bias. However, developmental studies mention fronting as a common error in young typically-developing children (as opposed to backing, which is common in atypical populations; see Bosch, 2004). In addition, children with CIs struggle to learn this feature (Bouton *et al.*, 2012). Given that speech recognition in noise is relatively difficult task for children (Johnson, 2000), it is possible that a fronting bias appears in children with NH (relative to adults with NH) and that the effect is increased in children with CIs, possibly due to a preference for visual information.

II. METHOD

A. Target consonant-vowels (CVs)

The speech targets for this study consisted of 80 CV syllables. The CV syllables were the exhaustive combination of

the 16 consonants that appear word-initially in Spanish and the five Spanish vowels (i.e., /a, e, i, o, u/). Table I shows the full list of consonants classified according to three phonological features. Based on data by Moreno-Torres *et al.* (2017) the 16 consonants included in this study can be divided into three resistance groups:

- Low resistance: /p, b, t, d, k, g, f, θ, x/
- Mid resistance: /m, n, l, r/
- High resistance: /s, ʃ, ʎ/

B. Background noise and stimuli

The stimuli used in this study are a subset of the items used in a previous study with adult listeners (Moreno-Torres *et al.*, 2017). The stimuli were created by combining the 80 CV targets produced by one adult male speaker with a background noise. The background noise (babble-8) was created by combining eight talker voices (four female, four male) recorded in a sound-proof room, reading Spanish language dialogs. The total duration of the babble-8 noise background was 1.2 s. The target CV began 300 ms after the babble-8 noise. The individual intensity levels for the babble-8 and target CVs were adjusted according to the global root mean square power of the original sounds to be mixed, at these SNRs: -6 dB, 0 dB, +6 dB, and +12 dB. Note, however, that not all the participants were evaluated at the same SNRs (see details below, in Sec. II E).

C. Listeners

The data for this study were obtained from a large database that includes the results of the experimental task described below. The database includes data from 26 children with CIs, 50 children with NH, and 23 adults with NH. The children with CIs were Spanish-speaking monolingual children wearing one ($N=23$) or two CIs ($N=3$). All the children with CIs had profound bilateral auditory loss detected pre-linguistically, and each of them had received at least one implant before their second birthday. While some children had used a hearing aid before implantation, post-implantation the only devices used were one or two CIs. The implant models used by the children were Cochlear ($N=16$), Advance Bionics ($N=7$) or Med-El ($N=3$). The mean age at implantation was 19 months (range: 11–30

months). The age of the children with CIs at evaluation ranged between 7 and 13 years ($M=9.4$). The mean time using the first CI was 7.9 years (range 6–12). The children had no impairments associated with auditory loss, although we were not able to conduct a formal evaluation of the language skills of the participants. However, all the CI users were enrolled in mainstream schools with either their age peers ($N=22$) or with children that were one year younger ($N=4$), which shows that they were able to obtain reasonably good results with their device. The ages of the children with NH ranged between 6 and 13 years ($M=9.1$). For the adult subjects with NH, the age ranged between 18 and 25 years ($M=20.1$). None of the NH subjects had previous history of auditory deficits or otitis. The parents of the children and the adult participants gave informed consent to participate in this study.

D. Testing conditions

The children with CIs in this study live in a relatively large area (southern Spain), and so it was not possible to evaluate them in a sound-proof room in our lab. For this reason, the children were evaluated in naturalistic context using standard loudspeakers (Gigaworks T20 Creative; SNR 80 dB; frequency response 50 Hz–20 kHz). In order to ensure that the conditions were similar for all the participants, a measure of the reverberation time (i.e., RT60) was obtained in each case. Previous studies have shown that this measure can be used as an estimation of the environmental conditions (e.g., Inglehart, 2016). For children and adults with NH, RT60 values below 0.60 have very limited impact on speech perception (e.g., speech recognition scores in rooms with these RT60 < 0.60 are similar to scores in quiet rooms). For children with CIs, the same effects occur when RT60 < 0.30. This means that levels of reverberation in the range 0.30–0.60 may pass undetected for children with NH but are disturbing for children with CIs.

In this study, the evaluations took place at RT60 values ranging between 0.26 and 0.52. When RT60 was larger than 0.52 in the participant’s home and there was not an alternative evaluation site, the participant was discarded from the study. The mean RT60 was 0.44 for the children with CIs, 0.42 for the children with NH, and 0.47 for the adults with NH. Given that CI users are more vulnerable to reverberation than NH subjects, these conditions may have increased the distance between the groups. However, it is important to note that reverberation is higher in large rooms (e.g., a typical school classroom) than in smaller ones (e.g., a typical living room). Thus, using a RT60 in the 0.30–0.60 range might provide a realistic measure of what these children hear in real-life situations.

E. Procedure

The listening test was automated using a Praat MFC Experiment code with graphic user interface (Boersma, 2001). The listener was seated in a comfortable chair in front of a computer monitor and heard the stimuli via the two loudspeakers. The monitor screen showed 17 buttons. Sixteen buttons were labeled with each of the consonants,

TABLE I. Consonant groups on the basis of voicing, manner and place of articulation.

Features	Values	Members
Manner	Plosive	p, t, k, b, d, g
	Affricate	ʧ
	Fricative	f, θ, s, ʃ, x
	Nasal	m, n
	Approximant	l, r
Place	Labial	p, b, f, m
	Coronal	θ, t, d, s, n, l, r, ʃ, ʧ
	Dorsal	k, g, x
Voicing	Voiced	b, d, g, ʃ, m, n, l, r
	Unvoiced	p, t, k, ʧ, f, θ, s, x

and one was unlabeled (i.e., “I do not know” button). As in previous studies in our lab, the listener was asked to use the mouse to select the response. As some of the participants were relatively young children, and the task requires increased attention effort, we decided that when necessary, the experimenter would manipulate the computer mouse throughout the task. In these cases, the child was instructed to imitate the syllable heard while looking at the researcher. When considered necessary by the researcher (~20% of the cases), clarification was requested by asking the child to point with their finger to the selected response.

The loudspeakers were placed in front of the listener at a distance of 80 cm, and the sound intensity was set to a level that ranged from around 64 to 72 dB sound pressure level (SPL). It is possible that this value produced further distortion of the speech signal for some of the children with CIs. However, once more we assumed that such distortion is not too different from the distortion that these children find in real-life situations.

Once the listener was familiar with the type of stimuli used, the evaluation began. The evaluation included the 80 CVs presented at three SNRs (i.e., 240 tokens). For the children with CIs, the SNRs were 0, +6, and +12 dB. For the children and adults with NH, the SNRs were -6, 0, +6 dB. In order to have a more diverse database so as to compare the NH and CI participants, one group of children with NH ($N=7$) were also evaluated at an SNR of +12 dB. Finally, in order to have a baseline measure of their recognition scores, all the children with CI and NH and one group of adults with NH ($N=7$) were evaluated without the loudspeaker. For this evaluation (Live condition), the researcher was placed 1 meter in front of the listener with a sheet of paper hiding his mouth. For each 80-syllable set, including the Live condition, a hits ratio was calculated.

F. Data selection

Based on the above-described database, we selected a total of 60 (i.e., 20 per group) 80-syllable sets for which the hit rate was approximately 33% (range 24%–45%). In total we selected 1600 responses per each group of participants, respectively, children with CIs, children with NH, and adults with NH. These datasets were used for the majority of the analyses in this study.

G. Data analysis

First, we computed the confusion matrices (CMs) for the three datasets. CMs list, for each spoken sound, the number of times that each response has been selected. Based on these CMs, we computed three error patterns: (1) correct; (2) omission (i.e., selecting “I do not know”); and (3) substitution (i.e., selecting one incorrect consonant). The CMs also provided information about false positives, which enabled us to compute consonant biases (e.g., preference for /f/ over /θ/).

In order to make group comparisons, we computed the ratio of errors per participant for several phonological features (e.g., for omission, place of articulation errors, etc.) When the test conditions were met, we used parametric tests to compare the groups; otherwise we used the non-parametric

alternatives. In order to explore the group bias towards specific feature values (e.g., voicing-devoicing, stopping-fricativization; fronting-backing), each response was given a different score. For voicing, the scores were +1 (devoicing error), 0 (correct), and -1 (voicing error). For manner of articulation, the scores were +1 (stopping), 0 (correct), and -1 (fricativization). For place of articulation, the scores were +3 (velar to frontal), +2 (velar to alveolar), and -3 (frontal to velar). This allowed us to obtain a bias measure for each participant, and to compare the three groups.

III. RESULTS

Before analyzing the three datasets, we computed the mean percentage of correctly recognized consonants in the full database (i.e., for the 26 children with CIs, 50 children with NH, and 23 adults with NH). The results showed that the adults with NH were more resistant to noise than children with NH, and that these were more resistant than children with CIs (Fig. 1). For instance, the SNR at which the children with CIs, children with NH, and adults with NH recognized 33% of the consonants were, respectively, +7.8 dB, -3 dB, and -6 dB. This means that adults with NH tolerate approximately 3 dB more noise than children with NH, and children with NH tolerate approximately 11 dB more noise than children with CIs.

In order to clarify if the task might be too difficult for the children with CIs, we examined how the error patterns changed throughout the tasks. Specifically, we compared the results for the first 80 tokens and for the remaining 160 tokens. The results revealed that in the three groups, there was a consistent reduction in the percentage of omissions (children with CIs: -0.09; children with NH: -0.12; adults with NH: -0.07). However, the groups differed in the how the hit rate and substitutions changed from the first to the second part. In the children with CIs, the hit rate remained stable and substitutions increased significantly (+0.6; $p=0.03$). In the two NH groups, the substitutions were stable and the hit rate increased significantly (children: +0.10, $p=0.003$; adults: +0.07, $p=0.02$). Altogether these results suggest that the task was not too demanding for the children with CIs (as evidenced by the stability in the hit rate), but that they are less efficient learners than the controls.

One potential limitation of the data from the children with CIs is that their errors might be unsystematic (in which case, the database might not provide reliable information about their speech processing skills). In order to rule out this possibility, we computed the mean number of errors that involved only one feature. The number of one-feature errors was higher in the children with CIs ($N=367$) than in the children with NH ($N=254$) or adults with NH ($N=280$). This indicates that the CI users succeeded in processing completely or partially a large number of target syllables, and confirms the usefulness of the database to explore their speech processing skills.

We also examined the precise errors produced by the CI users in the Live condition. Consonant confusions involved in most cases the place of articulation feature (>75%). These errors occurred mostly within the voiceless stops (p/t/k), the voiced stops (b/d/g), the three voiceless fricatives (f/θ/x), and,

less frequently, between the two nasals (m/n). Voicing errors were observed in 15% of the stops, with a bias towards voicing (66% / 33%). Manner errors were uncommon (<5%). In the NH participants the errors in the live condition were scarce (<2%) and involved mostly three syllable pairs (i.e., fe/θe, fi/θi and se/θe).

A. Phoneme error patterns

Tables II and III present the CMs for the children with CIs and the children with NH. The CM of the adults with NH is available as supplementary material.¹ Note that in the tables the consonants are grouped according to their resistance to noise (from less to more resistant) and their phonological group (see Table I). Based on these CMs we examined the omissions, substitutions, and false alarms (FAs). A one-way between-subjects analysis of variance (ANOVA) comparing the percentage of “I do not know” responses confirmed the existence of group differences [$F(2,57) = 12.106; p < 0.001$]. *Post hoc* comparison with the Tukey-b test showed that the mean percentage for the adults and children with NH (0.27 and 0.22, respectively) were significantly different from the mean for the children with CIs (0.11). Note that this means that the children with CIs produced more substitution errors than the two control groups, which might be associated with a difficulty segregating the target from the background noise.

Next, we analyzed the FAs. The children with CIs produced more FAs than the two NH groups for the three voiced stops (i.e., /b, d, g/) and also for the five most audible consonants (i.e., /l, r, j, s tʃ/). The accumulated number of FAs for these eight consonants were 510, 247, and 198, respectively, for the children with CIs, children with NH, and adults with NH. In contrast, the NH groups produced more FAs for the three voiceless stops and the velar voiceless fricative. The accumulated number of FAs were 91, 211, and 279,

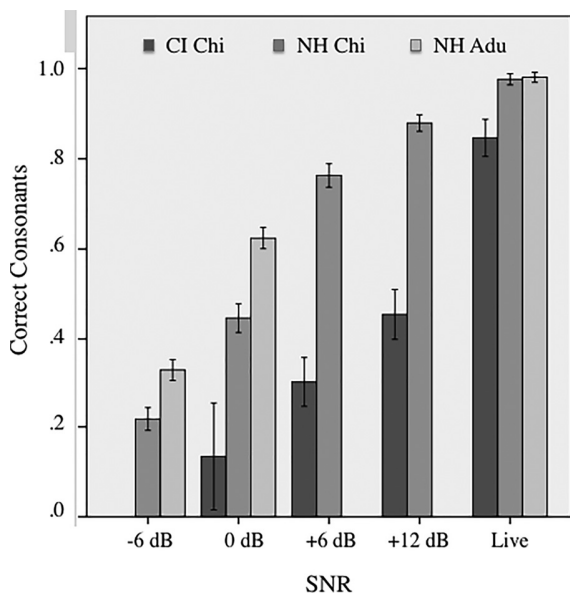


FIG. 1. Ratio of correct responses in the Live condition and with different SNR levels for the children with CI, the children with NH, and the adults with NH.

TABLE II. Confusion matrix for the children with CIs. The bold font indicates that the children with CIs selected the response nine or more times than the children with NH. FA: False Alarms (i.e., number of times a consonant is wrongly selected). The column * shows the number of “I do not know” responses per target consonant.

	Low resistance					Mid resistance											
	Stops			Fricatives		Nasals	Liquids	High resist.									
	*	p	b	t	d	k	g	f	θ	x	m	n	l	r	j	s	tʃ
p	23	10	13	8	13	4	7	9	6	2	2	0	0	1	1	1	0
b	14	1	34	4	12	1	8	7	2	1	9	0	3	2	2	0	0
t	14	0	6	25	10	9	9	2	9	0	2	1	0	3	2	2	6
d	8	0	9	1	38	0	9	1	5	0	7	5	8	2	6	0	1
k	9	1	5	6	10	22	16	4	14	1	1	0	3	4	1	1	2
g	7	0	11	1	20	5	41	0	1	0	1	1	5	6	1	0	0
f	12	3	8	1	6	0	2	22	31	2	1	3	0	0	2	6	1
θ	8	1	3	1	8	0	5	7	36	0	1	0	1	0	3	22	4
x	28	6	8	1	5	4	3	9	12	7	1	1	2	3	2	7	1
m	13	0	10	1	4	3	6	1	2	1	36	11	9	2	0	1	0
n	9	2	2	0	1	0	2	0	1	0	31	35	13	0	1	2	1
l	5	1	8	0	6	0	1	1	1	0	3	3	58	6	6	0	1
r	8	2	4	4	9	6	17	4	3	1	2	1	7	32	0	0	0
j	8	0	0	0	11	0	9	0	0	1	2	2	6	3	56	1	1
s	3	0	1	1	1	0	1	3	17	1	0	0	0	0	0	66	6
tʃ	4	1	1	1	1	0	0	3	5	1	0	0	1	0	4	21	57
FA	173	18	89	30	117	32	95	51	109	11	63	28	58	32	31	64	24

respectively for the children with CIs, children with NH, and adults with NH. In order to clarify the impact of audibility, we compared the hit rate for consonants with low, mid, and high audibility. The group difference was significant only for the most audible consonants (Kruskal-Wallis test $p = 0.003$), and only between the adults with and the

TABLE III. Confusion matrix for the children with NH. The bold font indicates that the children with selected the response nine or more times than the children with CIs. FA: False Alarms or the number of times a consonant is wrongly selected. The column * shows the number of “I do not know” responses per target consonant.

	Low resistance					Mid resistance											
	Stops			Fricatives		Nasals	Liquids	High resist.									
	*	p	b	t	d	k	g	f	θ	x	m	n	l	r	j	s	tʃ
p	36	5	8	8	11	5	3	7	5	2	1	0	1	0	1	7	0
b	33	3	19	2	5	7	2	6	5	2	9	2	3	0	0	1	1
t	23	0	2	41	7	8	0	2	10	1	0	2	1	0	1	2	0
d	30	2	1	15	31	5	0	0	5	0	1	1	6	1	1	1	0
k	30	7	2	7	5	23	9	1	3	1	2	2	5	2	1	0	0
g	23	1	4	0	9	28	24	3	1	1	2	0	2	1	1	0	0
f	21	9	2	1	1	3	3	25	28	3	1	0	1	0	0	2	0
θ	8	0	0	3	3	3	1	9	62	4	0	2	1	0	0	4	0
x	37	4	3	4	2	9	4	1	0	32	1	0	2	0	0	0	1
m	20	2	3	1	0	5	2	5	1	0	55	3	0	0	0	1	2
n	11	1	2	1	5	0	2	1	0	0	24	42	10	0	0	1	0
l	10	2	1	2	6	2	1	0	5	1	4	5	59	0	0	2	0
r	40	0	3	6	0	17	3	2	5	0	1	0	3	17	1	1	1
j	22	0	0	2	9	4	7	0	5	0	2	1	3	0	42	1	2
s	5	1	1	3	2	1	0	1	5	4	0	0	1	0	0	72	4
tʃ	8	0	0	2	3	3	3	2	7	3	1	2	0	1	5	18	42
FA	357	32	32	57	68	100	40	40	85	22	49	20	39	5	11	41	11

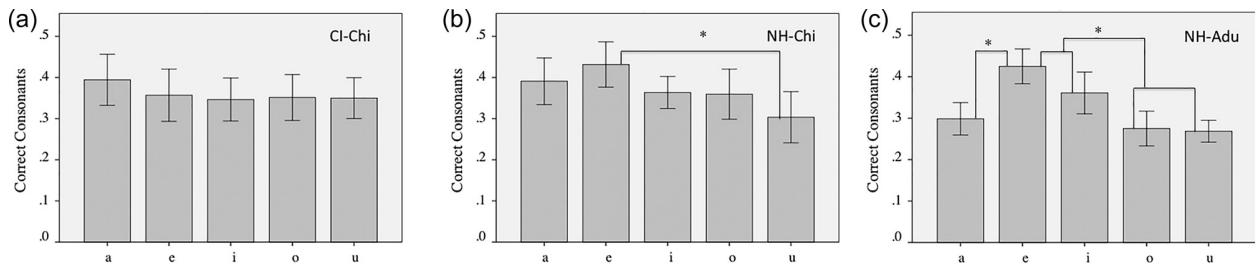


FIG. 2. Correctly recognized consonants as a function of the accompanying vowel for the children with CI, the children with NH, and the adults with NH. * $p < 0.05$.

children with CIs. Finally, we examined the results for the pair /f, θ/. In the three groups there was a bias for the more frequent consonant (i.e., /θ/). The two groups of children selected /θ/ between three and four times more frequently than /f/. The adults with NH selected the selected /θ/ eight times more frequently than /f/.

Altogether, these results show that the children with CIs produced errors that are different from those of the two NH groups: they produced more substitution errors than the controls, they were relatively successful with consonants that are highly audible, and they had a preference for the voiced stops. In contrast, the error patterns of the children and adults with NH were very similar to each other.

B. Errors related with the use of dynamic cues

Figure 2 shows the percentage of correct responses as a function of the accompanying vowel. In the children with CIs, the scores were almost identical for the five vowels. In the two NH groups, there were clear vowel effects. As the results did not have a normal distribution in some cases, we analyzed the vowel effect using the Kruskal-Wallis non-parametric test. The effect of vowel was significant in the children with NH ($p = 0.003$) and in the adults with NH ($p < 0.001$). Pairwise comparisons showed that the difference was significant in one pair in the children with NH, and in four cases in the adults with NH (see Fig. 2).

Figure 3 shows the mean ratio of feature errors for voice, place of articulation, and manner of articulation. For the manner and voice features, the groups were not different. For the place feature, the children with CIs produced more errors than the two NH groups ($p < 0.001$). Altogether these results indicate that, as predicted, consonant recognition in children with CIs does not depend on the accompanying vowel, and that they struggle to recognize the place of articulation feature. The results also indicate that vowel effects might be more pronounced in adults with NH than in children with NH.

C. Error biases

Bias was examined for the three features indicated above. For voicing, the children with CIs showed a clear voicing bias, the children with NH showed a balance between voicing and devoicing, and the adults with showed a slight devoicing bias (see Fig. 4). A one-way ANOVA with group as factor and voicing bias as the dependent variable confirmed a main effect of group [$F(2,57) = 13.697$;

$p < 0.001$]. *Post hoc* pair-wise comparisons using Tukey-b tests showed that there were significant differences between the children with CIs and the two NH groups. For the place of articulation bias (see Fig. 4), a one-way ANOVA again revealed a main effect of group [$F(2,57) = 11.187$; $p < 0.001$]. *Post hoc* pairwise analyses revealed the existence of significant differences between the adults with NH (backing bias) on the one hand, and the two groups of children (fronting bias) on the other. The fronting bias was somewhat stronger in the children with CIs than in the children with NH, but this difference did not reach significance. Finally, regarding the manner of articulation in the three groups, there was a balance between stopping and fricativization.

IV. DISCUSSION

The main aim of this study was to identify qualitative, not quantitative, differences in how children with CIs and children with NH process speech in noisy contexts. One difficulty in addressing this issue is that speech recognition in noise is notably more difficult for CI listeners than for NH listeners. This was confirmed by the results from our full database of 26 children with CIs and 73 subjects with NH. As Fig. 1 shows, when the hit rate is 33%, the group distances in SNR are 11 dB between children with CIs and children with NH, and 14 dB between children with CIs and adults with NH (a result similar to Chun *et al.*, 2015). This means

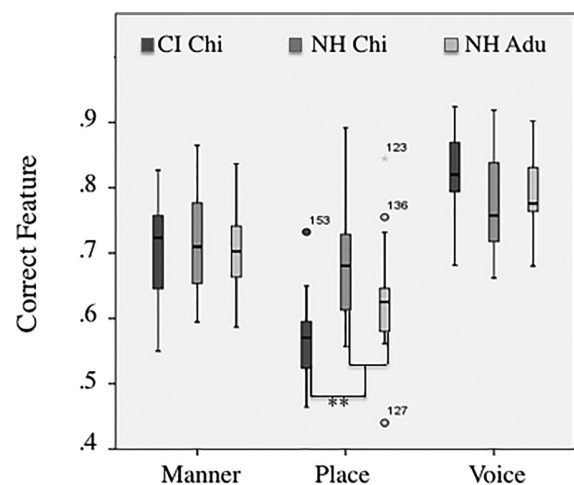


FIG. 3. Mean ratio of feature errors for the three groups (excluding “I do not know” responses) for the children with CI, the children with NH, and the adults with NH. The group comparison was made using Kruskal-Wallis test. ** $p < 0.001$.

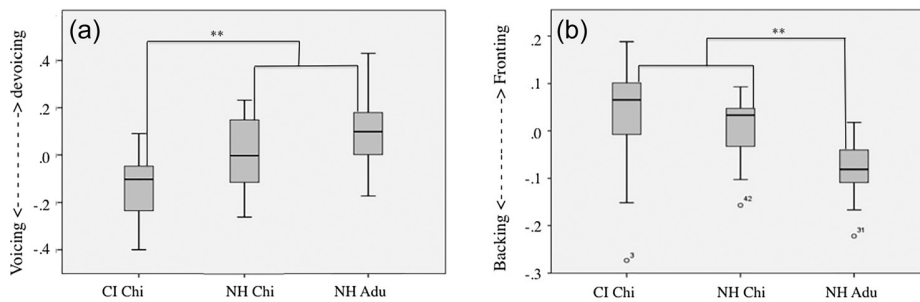


FIG. 4. Voicing (left) and fronting (right) bias for the children with CI, the children with NH and the adults with NH. $**p < 0.001$.

that if the groups of participants are matched on age exclusively, the errors produced by the CI users are so severe that it is not possible to analyze how they process speech, which shows the need to use a more stringent criteria to match the groups. In this study, the groups were matched on the ratio of correct responses (30%–33% correct).

One potential limitation of this study is that, due to the rapid signal degradation in noise, the children with CIs might have tended to produce random answers. However, our results showed that the three groups of participants produced similar percentages of one-feature errors. This suggests that, even if the children with CIs produced more substitution errors than the controls, in many of these cases they were still able to process part of the acoustic information, which shows that the task was appropriate to answer the questions addressed in this study.

A. Which phoneme error patterns do CI users produce?

The groups differed notably in the ratios of “I do not know” responses. The percentage was significantly higher in the NH groups than in the children with CIs. Furthermore, in the children with CIs the substitution errors increased from the first 80 tokens to the last 160 tokens in the task, while in the NH groups, the percentage remained stable. As for the consonant errors, two main differences emerged between the children with CIs and the two NH groups. One is that the children with CIs produced more FAs for the three voiced stops and for the most audible consonants than the controls. Another is that the controls produced more responses for the voiceless stops than the children with CIs. The three groups showed a similar preference for /θ/ over /f/, though the preference was more pronounced in the adults with NH than in the children.

Thus, as predicted, CI users produced a large number of substitution errors. These errors support our proposal that the difficulty segregating the target and noise signals (Qazi et al., 2013) and hence the amount of informational masking would make it difficult for children with CIs to disregard ambiguous input. This result is relevant because it suggests that children with CIs might have more difficulties than children with NH for language processing (i.e., noise does not only alter speech processing, it also alters language processing). Note that “I do not know” responses indicate that the listener may build incomplete phonological representation (i.e., with missing phonemes). In contrast, substitution errors may result in ill-formed representations (i.e., with wrong

phonemes). The consequences of each of these errors for later language processing might be very different. In the case of incomplete representations, the listener might be rapidly aware (i.e., before lexical processing) of the loss of information, for which they may either attempt to choose among plausible candidates (i.e., one of the few available phonological neighbors) or simply cancel lexical recognition. For instance, if listeners recognize the phoneme sequence /t_bl_/ (after hearing the English word “table”), the list of lexical candidates might include: “table,” “tubular,” etc. While it is possible that a wrong decision might be made, it is relevant that the listener might be aware of this possibility. In contrast, ill-formed representations may misguide the subject to make wrong lexical decisions, which may pass undetected to the listener. For instance, if under the same conditions (i.e., presented with the word “table”) a listener recognizes the phoneme sequence /maple/, the corresponding lexical word would be selected (i.e., “maple”), and the listener would not be aware of the lexical error. In the end, frequent ill-formed representations may make the listener unconfident of whatever interpretation is made of the speech signal, which may result in an increase in the cognitive resources needed to double-check every lexical recognition decision. Thus, the fact that children with CIs produce an increased number of substitution errors might have the most negative effects for language interaction in noisy contexts.

The fact that the children with CIs were especially successful with the most audible consonants confirms our prediction regarding these consonants: as these listeners have poorer auditory acuity, they may have better learned the sounds that have more intrinsic energy. This result is compatible with the results of a previous study including some of participants in this study showing an advantage for sibilant consonants in speech production (Moreno-Torres and Moruno-López, 2014). Thus, this result shows that audibility is a relevant factor when explaining speech and language development in CI users.

B. What factors may explain the errors observed in children with CIs and children with NH?

1. Dynamic acoustic cues

In the NH listeners, the ratio of recognition varied as a function of the accompanying vowel, an effect which was more pronounced in the adults than in the children. Specifically, in the children with NH, the scores were significantly lower with the vowel /u/ than with the vowel /e/. In the adults with NH, differences were significant between the

two back vowels (i.e., /o, u/) and between the two front vowels (i.e., /i, e/). From an acoustic perspective, these results are suggestive of a second formant (F_2) effect. Note that the second formant values are: $i, e > a > o, u$. Thus, the recognition scores increased as F_2 increased. Why did vowels with high F_2 facilitate consonant recognition? This effect is possibly related to the spectral structure of babble, which concentrates the largest amount of energy in the lower parts of the noise spectrum (see [Moreno-Torres et al., 2017](#)). This means that the listeners could benefit from the information provided by F_2 in the front vowels (i.e., those with highest F_2). In contrast, the CI users recognized the consonants similarly independently of the adjacent vowel, which indicates that they ignored or did not have access to the formant transitions. In sum, as predicted, the NH listeners, but not the CI users, used the vowel (i.e., the formant transitions) to categorize the preceding consonant. Finally, the fact that, despite the evidence that children tend to pay more attention than adults to vowel transitions ([Nittrouer, 2004](#)), the effect was more pronounced in the adults with NH than in the children with NH indicates that language experience may increase listeners' capacity to recognize formant transitions. Thus, our results seem to be compatible with the proposal that adults are more efficient than children in recognizing some complex acoustic sequences (i.e. [Mayo et al., 2016](#)). Note however that the lack of advantage for the children is because the participants in this study were older (from 6 to 13) than the participants in previous studies (from 5 to 7).

The results showed that the three groups of participants scored similarly for manner of articulation and voicing errors, but not for place of articulation errors. The results for manner of articulation are not surprising. They confirm that the children with CIs succeed in recognizing temporal cues (e.g., [Bouton et al., 2012](#)). The results for place of articulation are also compatible with the results of previous studies (e.g., [Tye-Murray et al., 1995](#)), and provide further evidence of the difficulties involved in recognizing dynamic spectral cues (e.g., [Hedrick et al., 2011](#)). It is also relevant to note two subtle but potentially relevant differences between the children and adults with NH. The children with NH were more successful, and more variable, than the adults with NH in recognizing place of articulation. These results seem to be compatible with these two proposals: (1) that young children are more efficient than adults in recognizing dynamic cues (e.g., [Nittrouer, 2004](#)), which is supported by the fact that children recognized place of articulation better than adults; and (2) that language experience continues to shape speech processing during adolescence ([Mayo et al., 2016](#)), which is supported by the fact that the children were more variable than the adults. Finally, the results in this study might be valuable in order to clarify the age at which children are perceptually mature. The participants in the studies by Nittrouer and colleagues were relatively young (5–7). Despite being somewhat older (6–13), the participants in this study were not identical to the adult participants, which indicates that developmental changes on how children attend to formant transitions might continue after the eighth birthday.

To summarize, as predicted, the children with CIs produced a large number of errors whenever they were required

to process dynamic acoustic cues. This result shows that, similarly to what has been observed in adults with CIs ([Munson and Nelson, 2005](#)), children with CIs struggle to process dynamic acoustic cues in noise as well. The results also suggest that while the differences between the adults and children with NH are limited (i.e., as compared with children with CIs) they are not identical: language experience continues to shape speech processing in NH listeners after the age of 7–13 (i.e. the ages of the children with NH in this study).

2. Error biases

The examination of the biases revealed that the groups were identical with regards to manner of articulation, but not for voicing or place of articulation. The children with CIs showed a clear voicing bias that contrasts with the devoicing bias observed in both NH groups. As for place of articulation, both groups of children showed a fronting bias as compared with the adults with NH.

To our knowledge the voicing bias observed in this study has not been described previously. However, it is important to note that previous studies have provided limited details about the errors observed in noise in CI users. In addition, this effect might be specific to languages in which voicing is associated with a short lag (making it acoustically weaker), such as Spanish, and not in languages with long lag, such as English. Thus, it is possible that this bias has passed undetected or it not been observed in previous studies because it is language-specific.

In order to explain this effect in NH listeners, it is important to note that babble noise more effectively masks the lower part of the spectrum (<1000 Hz), which may result in the sonority bar being energetically masked. When the sonority bar is masked, NH subjects tend to show a devoicing bias (e.g., [Gurlekian et al., 1987](#); [Moreno-Torres et al., 2017](#)). This bias might be caused by the combined effect of energetic masking, which masks the sonority bar, and phonological preference, which guides the listener towards the unmarked member in each pair.

And why are CI users biased in the opposite direction? One possible explanation is that their decisions are indirectly caused by the differences in audibility. That is, because voiced consonants are more audible, CI users may learn them better than the voiceless counterparts. While this explanation confirms our prediction, it is important to note that the contrast is very high. Indeed, in previous studies, audibility has been shown to produce subtle effects (e.g., [Moreno-Torres and Moruno-López, 2014](#)), while in the present study, the effect is very robust. This shows the need to consider another explanation. One possibility is that CI users might interpret masking noise as a sonority bar. In other words, they might be misguided by informational masking.

As for the stopping-fricativization bias, it was not present in any of the groups. This result is contrary to our prediction that informational masking would create the illusion of a nonexistent turbulent noise (which might produce a bias towards fricativization in the children with CIs). One potential explanation for this result is that the durational

cues block the fricative interpretation, and thus informational masking is avoided. In other words, as noise has a negative impact on a secondary cue, its effect is almost inexistent.

Finally, as regards the fronting-backing bias, we found that, as compared with the adults, both children with CIs and children with NH showed a preference for frontal consonants. This result indicates that during this period of development, children might have a more robust knowledge of visible consonants than they have of non-visible ones. This result is compatible with the evidence that very early in development, typical children tend to show fronting errors (Bosch, 2004). Note also that the fronting bias is slightly higher in the children with CIs than in the children with NH. While we did not find a statistical difference between the two groups of children, the results are compatible with the possibility that the poor quality of the input received by these children may result in an increase in the use of visual information to develop phonology.

To summarize, the results show that in the case of the voicing feature, the CI users showed an error bias that is clearly different from the one typically found in NH listeners and also in this study. This difference might be related to the technical limitations of their devices and/or to atypical phonological skills. The fact that both children with NH and children with CIs showed a slight fronting bias provides further evidence that speech processing is not mature in this age range (7–13).

C. Implications for future research

The present study has implications for future research. In the first place, while a clear result of this study was the voicing bias observed in the CI group, we could not clarify whether the bias was caused by phonetic factors (i.e., informative masking) or by phonological factors (i.e., having a more robust phonological knowledge of those speech sounds that are more energetic). One approach to clarify this issue would consist of analyzing how CI users interpret manipulated syllables in which the sonority bar has been removed (Feijóo *et al.*, 1988). If CI users show the same voicing bias as in this study, it would reinforce the proposal that audibility determines the phonological representations (i.e., the unmarked phonemes might be those with the highest audibility). Alternatively, if they show no bias, the results would reinforce the hypothesis that the errors are caused by informational masking. From a more general perspective, it might be most valuable to carry out similar experiments in other languages. Cross-linguistic data might be help to clarify the impact of phonetic and phonological aspects on how CI users process speech in noisy contexts.

The evidence presented here may also have implications for practical research in speech processing technology and speech rehabilitation. Regarding the first, to our knowledge the CI industry and research have not considered the possibility of designing speech processors specific to different types of spectral cues, particularly dynamic spectral cues. It might be useful to develop cue-specific processors that might parse different components of the speech signal separately; inasmuch as these cue-specific processors are successful, it

might be possible to emulate the cue weighting process observed in humans. Finally, from a clinical perspective, it might be useful to explore up to what point rehabilitation programs focused on the perception of dynamic spectral cues are beneficial for CI users. Note the fact the children with CIs paid more attention to the consonant than to the accompanying vowel, which may reflect a (badly) learned approach to speech processing. Thus, it is possible that this skill can be learned as part of speech therapy programs.

V. CONCLUSIONS

The most important results from this study can be summarized as follows:

- (1) The children with CIs recognized 33% of the nonsense syllables with a SNR of +7.8 dB. The children with NH obtained a similar score with a SNR of -3 dB, and the adults with NH with a SNR of -6 dB.
- (2) When matched on the hit rate, children with CIs produced more substitutions and fewer omissions than both children and adults with NH. This effect might be due to difficulty segregating the target signal from the background noise, which might result in increased informational masking.
- (3) As compared with NH listeners, children with CIs seem to be inefficient in using phonetic context, and they show increased difficulties with low-energy speech sounds. Like NH listeners, their processing skills improve with experience.
- (4) Children with NH show less pronounced context effects than adults with NH; the NH groups also differed in the fronting bias. These results suggest that speech-processing skills continue to be shaped at least during adolescence.
- (5) The fact that, when matched on hit rate, the children with CIs produce more severe errors than the children with NH shows that there are qualitative differences, not only quantitative, in how these two groups process speech in noise.

ACKNOWLEDGMENTS

This study was funded by the project DENSIC (El Desarrollo del Niño Sordo con Implante Coclear) awarded by the Spanish Ministerio de Economía y Competitividad (FFI2015-68498-P MINECO/FEDER) to the first author.

¹See supplementary material at <https://doi.org/10.1121/1.5044416> for the confusion matrix of the adults with NH.

- Albalá, M. J., and Marrero, V. (1995). "La intensidad de los sonidos españoles" ("The intensity of Spanish sounds"), *Revista de Filología Española* **85**, 1–15.
- Alexander, J. M., and Kluender, K. R. (2008). "Spectral tilt change in stop consonant perception," *J. Acoust. Soc. Am.* **123**(1), 386–396.
- Allen, P., and Wightman, F. (1992). "Spectral pattern-discrimination by children," *J. Speech Lang. Hear. Res.* **35**(1), 222–233.
- Boersma, P. (2001). "PRAAT, a system for doing phonetics by computer," *Glott Int.* **5**, 341–345.
- Bosch, L. (2004). *Evaluación Fonológica del Habla Infantil (Phonological Evaluation of Children's Speech)* (Masson, Barcelona, Spain).

- Bouton, S., Serniclaes, W., Bertonicini, J., and Cole, P. (2012). "Perception of speech features by French-speaking children with Cochlear Implants," *J. Speech Lang. Hear. Res.* **55**, 139–153.
- Brungart, D. S. (2001). "Informational and energetic masking effects in the perception of two simultaneous talkers," *J. Acoust. Soc. Am.* **109**, 1101–1109.
- Caldwell, A., and Nittrouer, S. (2013). "Speech perception in noise by children with cochlear implants," *J. Speech Lang. Hear. Res.* **56**(1), 13–30.
- Chun, H., Ma, S., Han, W., and Chun, Y. (2015). "Error patterns analysis of hearing aid and cochlear implant users as a function of noise," *J. Audiol. Otol.* **19**(3), 144–153.
- Dubno, J. R., and Levitt, H. (1981). "Predicting consonant confusions from acoustic analysis," *J. Acoust. Soc. Am.* **69**(1), 249–261.
- Feijóo, S., Fernández, S., and Balsa, R. (1998). "Integration of acoustic cues in Spanish voiced stops," in *Proceedings of the ICA/ASA Meeting*, June 20–26, Seattle, WA, pp. 2933–2934.
- Gerrits, E. (2001). "The categorisation of speech sounds by adults and children: A study of the categorial perception hypothesis and the developmental weighting of acoustic speech cues," Ph.D. thesis, Utrecht University, Utrecht, the Netherlands.
- Giezen, M. R., Escudero, P., and Baker, A. (2010). "Use of acoustic cues by children with cochlear implants," *J. Speech Lang. Hear. Res.* **53**(6), 1440–1457.
- Gurlekian, J. A., Guirao, M., and Franco, H. (1987). "On the identification of Spanish voiced stops," *J. Acoust. Soc. Am.* **82**, S119.
- Hawkins, S. (2010). "Phonological features, auditory objects, and illusions," *J. Phon.* **38**, 1, 60–89.
- Hazan, V., and Barrett, S. (2000). "The development of phonemic categorization in children aged 6–12," *J. Phon.* **28**, 377–396.
- Hedrick, M., Bahng, J., von Hapsburg, D., and Younger, M. S. (2011). "Weighting of cues for fricative place of articulation perception by children wearing cochlear implants," *Int. J. Audiol.* **50**(8), 540–547.
- Hnath-Chisolm, T. E., Laipply, E., and Boothroyd, A. (1998). "Age-related changes on a children's test of sensory-level speech perception capacity," *J. Speech Lang. Hear. Res.* **41**(1), 94–106.
- Inglehart, F. (2016). "Speech perception in classroom acoustics by children with cochlear implants and with typical hearing," *Am. J. Audiol.* **25**, 100–109.
- Johnson, C. E. (2000). "Children's phoneme identification in reverberation and noise," *J. Speech Lang. Hear. Res.* **43**, 144–157.
- Mayo, C., and Turk, A. (2004). "Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions," *J. Acoust. Soc. Am.* **115**(6), 3184–3194.
- Mayo, C., Turk, A., and Clark, R. (2016). "Predictability and adult-child cue weighting differences in speech perception," in *Proceedings of Speech Prosody 2016*, May 31–June 3, Boston, MA, pp. 553–557.
- McGurk, H., and MacDonald, J. (1976). "Hearing lips and seeing voices," *Nature* **264**, 746–748.
- Miller, G. A., and Nicely, P. (1955). "An analysis of perceptual confusions among some English consonants," *J. Acoust. Soc. Am.* **27**, 338–352.
- Moreno-Torres, I., and Moruno-López, E. (2014). "Segmental and suprasegmental errors in Spanish learning cochlear implant users: Neurolinguistic interpretation," *J. Neurolinguist.* **31**, 1–16.
- Moreno-Torres, I., Otero, P., Luna-Ramírez, S., and Garayzábal-Heinze, E. (2017). "Analysis of Spanish consonant recognition in 8-talker babble," *J. Acoust. Soc. Am.* **141**(5), 3079–3090.
- Munson, B., and Nelson, P. B. (2005). "Phonetic identification in quiet and in noise by listeners with cochlear implants," *J. Acoust. Soc. Am.* **118**(4), 2607–2617.
- Nittrouer, S. (2004). "The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults," *J. Acoust. Soc. Am.* **115**(4), 1777–1790.
- Nittrouer, S., Caldwell-Tarr, A., Moberly, A. C., and Lowenstein, J. H. (2014). "Perceptual weighting strategies of children with cochlear implants and normal hearing," *J. Commun. Disord.* **52**, 111–133.
- Nittrouer, S., and Crowther, C. S. (1998). "Examining the role of auditory sensitivity in the developmental weighting shift," *J. Speech Lang. Hear. Res.* **41**(4), 809–818.
- Qazi, O. U., van Dijk, B., Moonen, M., and Wooters, B. (2013). "Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility," *Hear. Res.* **299**, 79–87.
- Simpson, S. A., and Cooke, M. (2005). "Consonant identification in N-talker babble is a nonmonotonic function of N_s ," *J. Acoust. Soc. Am.* **118**, 2775–2778.
- Sroka, J., and Braidá, L. D. (2005). "Human and machine consonant recognition," *Speech Commun.* **45**, 401–423.
- Sussman, J. E. (2001). "Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions?," *J. Acoust. Soc. Am.* **109**, 1173–1180.
- Tye-Murray, N., Spencer, L., and Woodworth, G. G. (1995). "Acquisition of speech by children who have prolonged cochlear implant experience," *J. Speech Lang. Hear. Res.* **38**, 327–337.
- van Zyl, M., and Hanekom, J. J. (2013). "Perception of vowels and prosody by cochlear implant recipients in noise," *J. Comm. Disord.* **46**(5–6), 449–464.
- Wang, M. D., and Bilger, R. C. (1973). "Consonant confusions in noise: A study of perceptual features," *J. Acoust. Soc. Am.* **54**(5), 1248–1266.
- Woods, D. L., Yund, E. W., Herron, T. J., and Ua Cruadhlaioich, M. A. I. (2010). "Consonant identification in consonant-vowel-consonant syllables in speech spectrum noise," *J. Acoust. Soc. Am.* **127**, 1609–1623.