Otology & Neurotology
40:e191–e197 © 2019, Otology & Neurotology, Inc.

Assessing Fine-Grained Speech Discrimination in Young Children With Bilateral Cochlear Implants

Zhao Ellen Peng, Christi Hess, Jenny R. Saffran, Jan R. Edwards, and Ruth Y. Litovsky

Waisman Center, University of Wisconsin-Madison, Madison, Wisconsin

Hypothesis: Children of 2 to 3 years old with cochlear implants can perform consonant discriminations using fine-grained acoustic cues.
Background: Children born with severe-to-profound deafness are provided with early cochlear implantation (<2 yr) to maximize oral communication outcomes. Little is known regarding their abilities to discriminate consonant contrasts for accurately identifying speech sounds.
Methods: Using a Reaching for Sound paradigm to collect behavioral responses, consonant contrast discrimination was measured in 13 children with bilateral cochlear implants (BiCIs; aged 28–37 mo), and 13 age-matched normal-hearing (NH) children. Four contrast pairs were tested: 1) place + voicing, 2) place, 3) voicing, and 4) reduced voice-onset-time cue. Using standard processing strategies, electrodograms showing pulsatile stimulation patterns were created retrospectively to assess the spectral-temporal cues delivered through the clinical speech processors.
Results: As a group, children with BiCIs were able to discriminate all consonant contrasts at a level that was above chance, but their performance was poorer than NH children. Larger individual variability in discrimination performance was found in children with BiCIs. Stepwise regression revealed that, in the place contrast, chronological age was correlated with improved discrimination performance among children with BiCIs.
Conclusion: Children with BiCIs were able to discriminate consonant contrasts using fine-grained spectral-temporal cues above chance level but more poorly than their NH peers. Electrodogram analysis confirmed the access to spectral-temporal cues in the consonant contrasts through clinical speech processors. However, the cue saliency might not have be enough for children with BiCIs to achieve the same discrimination accuracy as NH children. **Key Words:** Bilateral cochlear implants—Children with cochlear implants—Consonant discrimination.

Otol Neurotol **40:**e191–e197, 2019.

Children who were born deaf or acquired severe to profound hearing loss soon after birth can (re)gain access to auditory information through cochlear implants. The use of cochlear implants has been shown to improve speech perception in school-aged children with profound hearing loss (1–5). With shifting criteria in medical treatment, the age at which cochlear implants are being provided to young children has decreased dramatically. Many children are now implanted before 2 years of age, and with bilateral implantation, with the goal of maximizing spoken language acquisition and oral communication (4).

In the area of speech perception, consonant discrimination is particularly important for accurate word understanding. By 12 months of age or even earlier, typically developing infants with normal hearing (NH) are able to accurately discriminate consonant contrasts with fine spectral-temporal distinctions (6–9). Children with cochlear implants tested between 5 and 7 years of age are better at differentiating consonant contrasts based on voicing (e.g., /b/ versus /p/), relative to consonant contrasts based on place of articulation (e.g., /p/ versus /t/). This is because children with implants are able to discriminate fine-grained temporal cues (i.e., small changes to voice onset time) at a level that is similar to their peers with NH (10,11). However, these children with implants perform more poorly for place of articulation contrasts, which require access to fine-grained spectral cues (10). This is likely due to poorer spectral resolution available through electrical hearing from a variety of factors, such as smaller range of audible frequencies and small number of frequency channels (12). There is also broader tuning of auditory neurons stimulated, resulted from wider spread of excitation with electrical stimulation as compared with the narrow tuning to acoustic stimulation (13,14). To better understand factors that may contribute to delays in speech recognition in young cochlear implant users, the current study was designed to investigate

Address correspondence and reprint requests to Zhao Ellen Peng, Ph.D., 1500 Highland Avenue, Madison, WI 53705; E-mail: z.ellen.peng@wisc.edu
The work was supported by grants from NIH-NIDCD (R01DC008365 to RYL, R01DC02932 to JE, R37HD037466 to JRS), and a core grant to the Waisman Center from NIH-NICHD (HD003352).
The authors report no conflicts of interest.
DOI: 10.1097/MAO.0000000000002115

consonant discrimination in 2- to 3-year-old children with cochlear implants and those with NH.

Previous behavioral studies used head-turn or preferential looking procedures to collect responses from young children. These methods require a large sample size, which is problematic for studies of low-incidence populations such as children with cochlear implants. They also do not readily permit assessment of multiple phoneme contrasts within a single study. To address these issues, we developed a novel ''Reaching for Sound'' method to capture perceptual discrimination. The task was modeled after studies with infants that measured reaching for sounding objects in the dark, in that it captures the natural behavior to reach for objects of interest (15,16). The reaching method was based on the approach used by Litvosky et al. (17) to study spatial hearing in young children, which engages their attention during a longer testing period to allow multiple trial repetitions at each condition.

MATERIALS AND METHODS

Participants

Two groups of children were tested: 13 children with cochlear implants and 13 NH children. All children in the implant group had bilateral cochlear implants (BiCIs), and had been previously recruited for a study on spatial hearing. All children with BiCIs had a history of severe-to-profound sensorineural hearing loss before they were implanted. At the time of testing, their chronological ages ranged from 28 to 37 months ($M = 32.5$ mo); they all had at least 12 months of listening experience with their first cochlear implant and used primarily auditory-verbal communication in English. Individual demographics and implant history are included in Table 1. During testing, the everyday program in their clinical speech processors was activated based on parental report.

Children in the NH group had no known developmental or neurological disorders, based on parental report. On the day of testing, none had ear infections, known illnesses or had taken medication, as reported by the accompanying parent or guardian. These individuals were sex-matched (with one exception) and age-matched (within ±2 mo chronological age) to individuals with BiCIs.

Experimental Materials and Tasks

Testing was conducted in a standard IAC sound booth (2.7 m × 3.6 m); children sat facing a semi-circular apparatus (1.5 m radius). A loudspeaker used to play sounds was placed at 0 degree azimuth, hidden behind a vertical curtain hung in front of the table. The curtain had two cut-out holes at +45 degrees and –45 degrees azimuth under the table, large enough for a small child to be able to reach their hand through and remove a small toy, to pose as the two-alternative forced-choice response options. Audio signals were prerecorded, calibrated, and played back (Tucker-Davis Technologies System 3) through a loudspeaker at 0 degree azimuth at ear level. A carrier phrase ''I'm hiding under'' was spoken by a female voice, followed by the target stimulus.

Target stimuli were chosen to capture different voice and place cues in consonants: /p/ voiceless bilabial plosive, /b/ voiced bilabial plosive, and /k/ voiceless velar plosive. Three core words ''bee,'' ''pea,'' and ''key'' were chosen. The words were recorded by the same female speaker of native English and processed at 44.1 kHz sampling rate. Two additional stimuli with modified voice onset time (VOT) were created using the recorded words ''bee'' and ''pea'' by digitally removing all burst release or aspiration based on methods described in Coady et al. (18). The final VOT value was 15 ms for the modified ''bee'' and 35 ms for the modified ''pea.'' The manipulation on VOT values on ''bee'' and ''pea'' brought the /p/-/b/ contrast closer to their phonetic boundary and increased discrimination difficulty. Target words were normalized in Boersma (19) to have equal root-mean-square energy and presented at 60 dBA sound pressure level (re 20 μPa) during testing. Four consonant contrasts were tested: 1) place + voicing [''bee'' versus ''key''], 2) place [''pea'' versus ''key''], 3) voicing [''bee'' versus ''pea''], and 4) reduced VOT [modified ''bee'' versus modified ''pea'']. Each of the three core words had an age-appropriate, clip-art style image, and appeared in pairs above

TABLE 1. *Subject demographics*

Subject Code	Sex	Chronological Age (mo)	Subject Code	Sex	Chronological Age (mo)								
Children With Normal Hearing			Children With Bilateral Cochlear Implants			Etiology	Age of 1st CI (mo)	Age of 2nd CI (mo)	Hearing Age (mo)	Bilateral Experience (mo)	Device Manufacturer; Processor	Maternal Education (yr)	Frequency of Therapy /Week (h)
CSF	M	26	CIFQ	M	28	Unknown	7	13	21	15	Cochlear; Nucleus 5	14.5	0.5
CRF	M	28	CIFZ	M	30	Connexin	8	8	22	22	Cochlear; Nucleus 5	16	3
CSB	M	29	CIFJ	M	30	Connexin	14	14	16	16	Med-EL; OPUS-2	17	1
CRE	M	30	CIFK	M	30	Connexin	14	14	16	16	Med-EL; OPUS-2	17	1
CRJ	M	32	CIFI	M	32	Unknown	7	7	25	25	Cochlear; Nucleus 5	20+	6
CRD	F	31	CIGB	F	32	Connexin	14	15	18	17	AB; Neptune	17	5
CQN	F	32	CIFY	F	33	Unknown	13	17	20	16	Med-EL; OPUS-2	17	4
CRU	M	34	CIFX	M	33	Unknown	14	15	19	18	Med-EL; OPUS-2	16	1
CRG	F	34	CIFO	F	34	Unknown	9	15	25	19	Med-EL; OPUS-2	16	2.5
CRW	M	35	CIFN	M	34	Unknown	13	13	21	21	Med-EL; OPUS-2	20	8
CSG	F	34	CIGA	F	34	Unknown	21	25	13	9	Med-EL; OPUS-2	14	1
CTM	M	35	CIFU	F	35	Connexin	12	15	23	20	Med-EL; OPUS-2	20	1.25
CRB	M	36	CIFT	M	37	Connexin	8	8	29	29	Cochlear; Nucleus 5	13	2.5
Mean =		32.0	Mean =		32.5		11.8	13.8	20.6	18.7		16.5	2.8

the two cut-out holes to represent the consonant contrast for children to choose as response option.

Before testing, each child underwent a familiarization procedure that lasted approximately 5 to 15 minutes using images of the target words that were later used during testing. If the child could correctly identify each of the target image by responding to the experimenter prompt ''Show me the ___,'' a brief puppet show was then used to introduce the task with the leading phrase ''I'm hiding under ___.'' The reinforcing puppet (or toy) was then hidden behind the curtain and the child was instructed to reach through the curtain to find it.

Experimental Procedure

The experimental protocol was approved by the Institutional Review Board at the University of Wisconsin-Madison. All parents provided written consent before children's participation in the study.

Similar to the procedure described in Litovsky et al. (17), three experimenters administered the test session. Experimenter 1 performed familiarization with the child, and sat with the child to provide necessary repositioning and reinforcement between trials. At the beginning of each trial, Experimenter 2 hid behind the curtain and initialized the trial by positioning a small puppet or toy above the center loudspeaker to capture the child's attention. Once the child was facing forward, the leading phrase ''I'm hiding under __'' followed by the target word was played from the loudspeaker. Experimenter 2 then removed the reinforcer from the child's view and placed it behind the hole at +45 degrees or −45 degrees on the side of the correct visual image and awaited the child's reach. The child responded by reaching for the hole under the image representing the word heard. The child's response to the stimulus was judged at the time of their reaching for the hole by Experimenter 3, who was blind to the audio stimuli and sat outside of the booth to perform real-time behavioral coding based only on visual observation.

Four blocks of test trials were initiated after the familiarization. Each block consisted of one contrast and ended after 16 valid trials. An invalid trial was defined as one in which the child refused to participate or reach after stimulus onset. On average, children in the NH group had fewer invalid trials ($M = 1.3$, SD = 1.7, maximum = 7) than children with BiCIs ($M = 2.6$, SD = 2.6, maximum = 11) in completing testing for each consonant contrast condition. The order of contrast block presentation and visual image location were randomized across children.

RESULTS

Percent correct scores of the consonant contrast discrimination were calculated based on 16 valid trials in each of the four test blocks, and converted to rationalized arcsine units (RAU) for statistical analyses (20) in SPSS (Version 22, IBM, Armonk, NY). Using the transformation, a perfect score of 100% is equivalent to 112 RAU, whereas a chance score of 50% is 50 RAU. Due to complications in the testing session, two test blocks were terminated at 11 and 12 trials for two children with BiCIs; this consisted of less than 2% of the overall dataset.

Fine-grained Speech Discrimination Performance

Accuracy in consonant contrast discrimination is reported in Figure 1 for all four contrast conditions for children with BiCIs and NH. Group-wise, children in both groups performed at a level above chance (50 RAU) in all contrast conditions ($p < 0.05$ through one-sample t tests). To assess differences in performance between children in the BiCI and NH groups, a two-tailed independent t test was conducted for each consonant contrast. Compared with their NH peers, children with BiCIs demonstrated poorer performance in accurately discriminating consonants in all four contrast conditions: place + voicing ($t[24] = 5.68$, $p < 0.001$); place ($t[24] = 4.64$, $p < 0.001$); voicing ($t[24] = 3.87$, $p = 0.001$); and reduced VOT ($t[24] = 2.68$, $p = 0.017$).

Within each child group, planned comparisons were conducted with Bonferroni corrections applied to a set of four paired t tests to examine the discrimination accuracy among different contrasts: 1) place + voicing versus place, 2) place + voicing versus voicing, 3) place versus voicing, and 4) voicing versus reduced VOT. Children with BiCIs did not show significantly different performance in any of the paired comparisons ($p > 0.05$). For NH children, discrimination performance was significantly better for place + voicing (/b/-/k/) versus voicing (/b/-/p/) ($t[12] = 3.86$, $p = 0.008$). Other comparisons were not significant. By modifying the VOT of /b/ and /p/ to reduce the voicing cue saliency, NH children's

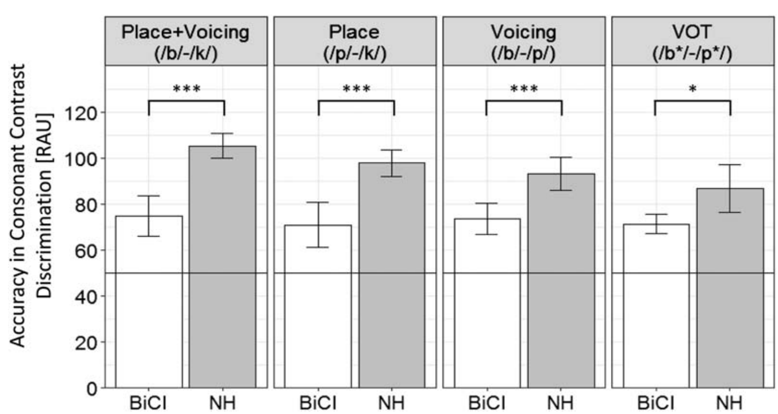

FIG. 1. Mean discrimination accuracy for children with BiCIs and in NH children in the four consonant contrast conditions. Error bars indicate 95% confidence interval of the mean reported. Asterisks indicate consonants with modified voice onset time. Solid line in each pane indicates chance level at 50 RAU. BiCI indicates bilateral cochlear implants; NH, normal-hearing; RAU, rationalized arcsine units.

averaged accuracy reduced only slightly (6 RAUs, equivalent to 6%) and this manipulation was not significant ($p > 0.05$).

Predictors of Speech Discrimination Performance

Individual data are plotted in Figure 2 to show discrimination accuracy for each child as a function of their chronological age. The spread of data points along the vertical axis was larger for children with BiCIs than for NH children. In the contrasts without VOT modification, individual children with BiCIs demonstrated a range of accuracy from levels that were close to chance performance at 50 RAU to perfect scores of 112 RAU (e.g., subject CIFT correctly identified all 16 trials in the /p/-/k/ place contrast). On the other hand, children in the NH group scored between 80 and 112 RAU with more individuals achieving perfect accuracy in consonant contrasts with higher cue saliency, for example, the place + voicing contrast.

To capture the possible predictors of variability in speech discrimination performance among children with BiCIs, an exploratory stepwise regression model was established for each contrast discrimination using a list of demographic variables. The demographic variables entered into the stepwise regression included chronological age, hearing experience (chronological age minus age at first implantation), bilateral cochlear implant experience (chronological age minus age at a second implantation), maternal education, and frequency of therapy per week. Chronological age in months was the only significant predictor of discrimination accuracy in the /p/-/k/ place contrast ($b = 4.66$, $t[11] = 2.79$, $p = 0.017$). It explained a significant proportion of variance in children with BiCIs for discrimination performance in the place contrast ($R^2 = 0.42$, $F[1, 11] = 7.80$, $p = 0.017$). For other consonant contrasts, however, none of the demographic variables were significant predictors in the stepwise regression models and were not correlated with the discrimination performance among children with BiCIs.

Access to Acoustic Cues in Cochlear Implant Processors

Accurate consonant identification and discrimination rely on fine spectral-temporal cues, as seen in the spectrograms of the five word tokens used in this study in Figure 3 (second column). To further understand the access to these spectral-temporal cues in children with BiCIs, pulsatile stimulation patterns in electrodograms for the five word tokens were

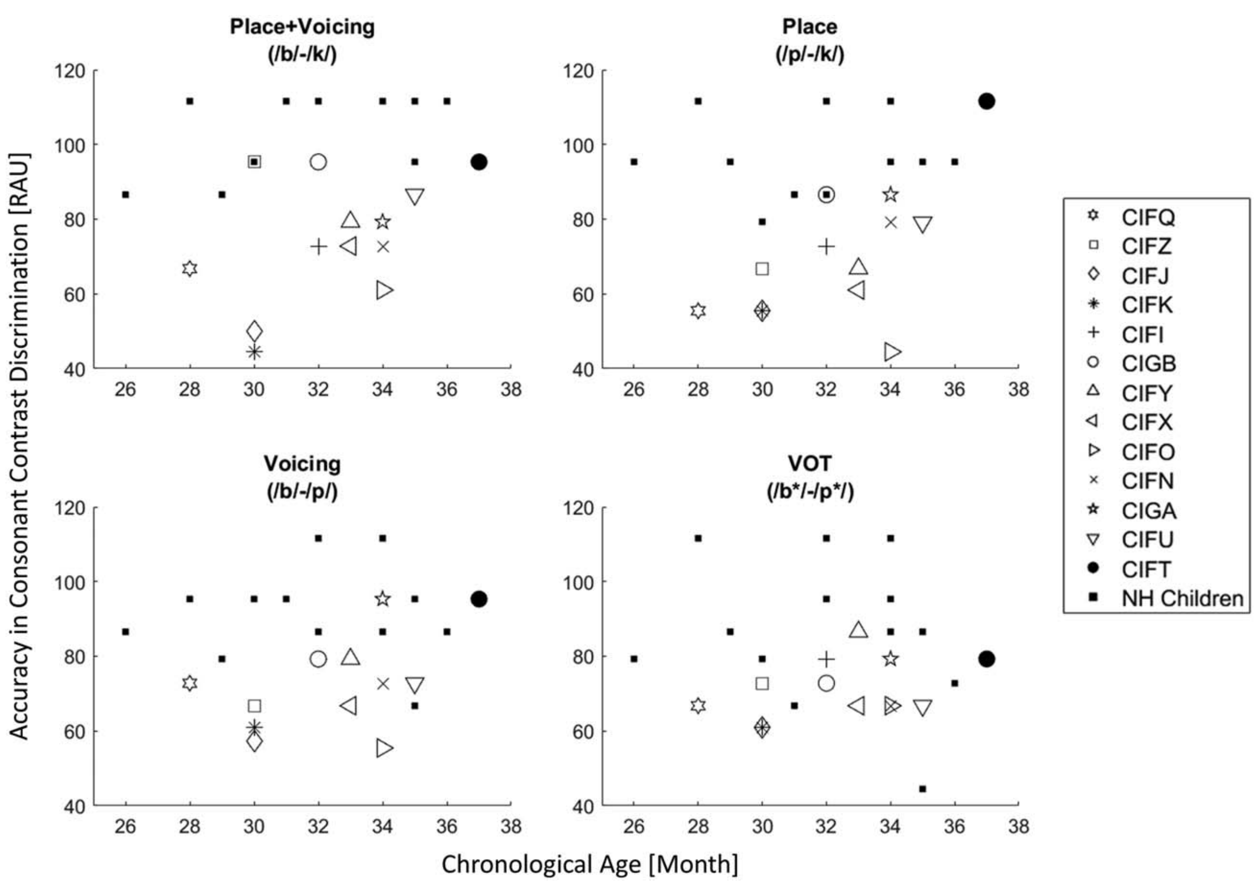

FIG. 2. Individual child's accuracy in consonant contrast discrimination plotted against chronological age at the time of testing. Individual data for children with NH in close squares. NH indicates normal-hearing.

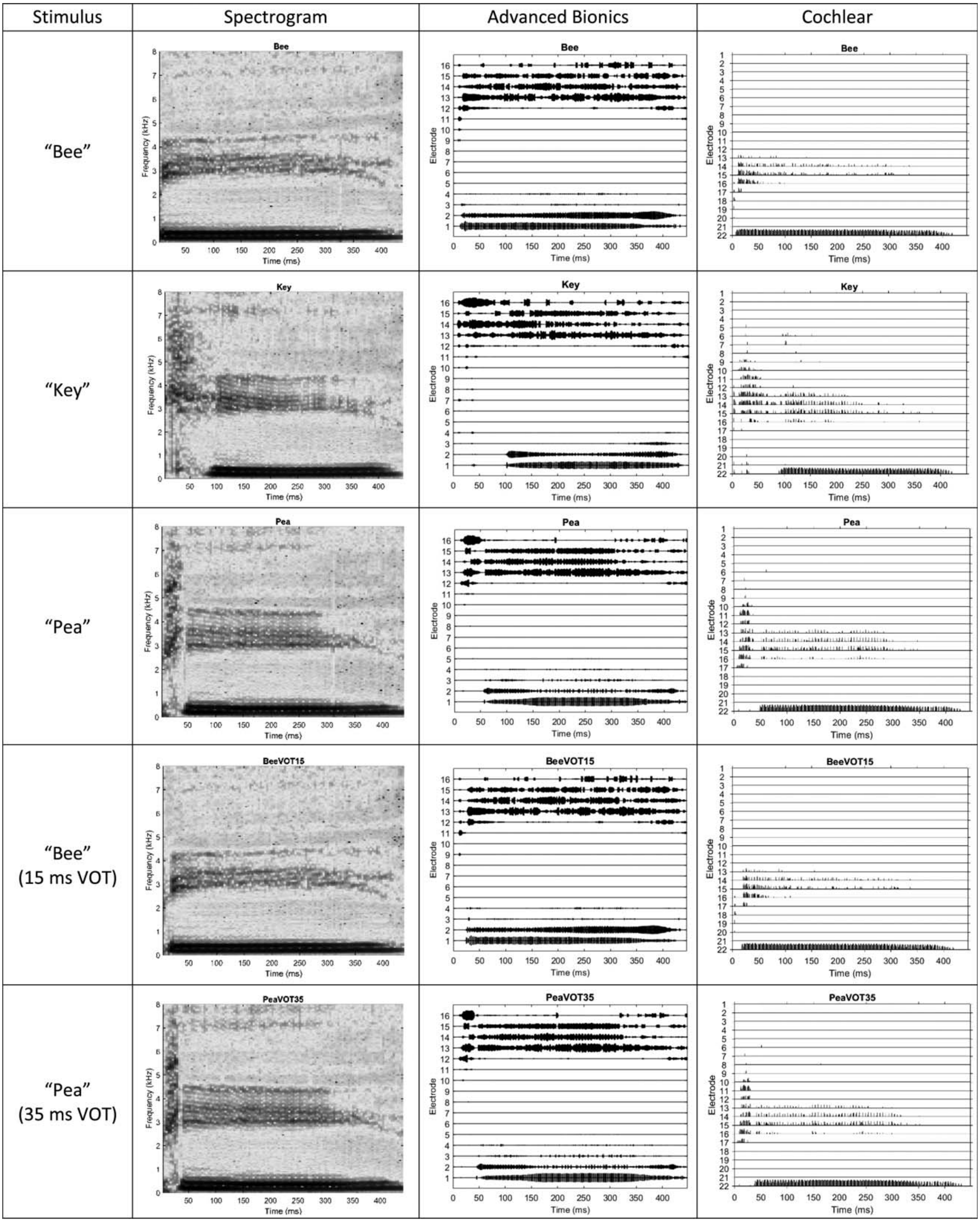

FIG. 3. Spectrograms of the five word tokens used in this study (second column). Simulated electrodograms showing pulsatile stimulation patterns for the HiResolution F120 Sequential (HiRes F120S) processing strategy from Advanced Bionics (third column) and the advanced combinational encoder (ACE) processing strategy from Cochlear (fourth column). All electrodograms were simulated using standard clinical maps. Electrode arrangement in the electrodograms was based on ascending frequency from bottom to top of graph instead of the order of electrode numbers.

simulated retrospectively using standard processing strategies and clinical maps for the devices used in this study. The simulated electrodograms were shown side by side with the spectrograms for individual target words: 1) for Advanced Bionics (AB) Neptune device using a 16-electrode standard clinical map with the HiResolution F120 Sequential (HiRes F120S) processing strategy (Fig. 3, third column), and 2) for Cochlear Nucleus device using a 22-electrode standard clinical map with the advanced combinational encoder (ACE) strategy (Fig. 3, fourth column).

As seen in the spectrograms, most of the acoustic energy that represented the consonants was concentrated between 2.5 and 5 kHz. From the electrodograms, both devices were able to stimulate at least five electrodes (electrodes 12–16 in HiRes F120S, electrodes 10–17 in ACE) to encode the acoustic signal above 3 kHz. In the first 100 ms that contained the consonant in each token, differing stimulation patterns both in intensity and duration were observed across electrodes. Although still largely coarse, children with BiCIs in this study did have access to spectral distinctions from their processors that represented the consonant contrasts.

DISCUSSION

Using the ''Reach for Sound'' paradigm to collect behavioral responses, fine-grained consonant contrast discrimination was studied in 13 children with BiCIs and 13 age-matched children with NH between 26 and 37 months old. Four consonant contrasts were tested: 1) place + voicing (''bee'' versus ''key''), 2) place (''pea'' versus ''key''), 3) voicing (''bee'' versus ''pea''), and 4) VOT (modified ''bee'' with 15 ms VOT versus modified ''pea'' with 35 ms VOT). When considering discrimination of consonant contrast that required access to fine spectral distinction, children with BiCIs achieved discrimination at a level that was better than chance; however, they performed worse than NH children.

One explanation for these results resides in the design of multi-channel cochlear implant which delivers poorer spectral resolution of consonant contrasts as compared with a healthy auditory system. Retrospective electrodogram analysis confirmed that children with BiCIs had access to spectral-temporal changes that represented respective consonant contrasts. But the spectral resolution was much poorer than those perceived by their NH peers and thus partially explained the poorer discrimination accuracy among children with BiCIs. From the simulated electrodograms of ''pea'' (/p/) and ''key'' (/k/), stimulation patterns differed across electrodes carrying high-frequency information within the first 100 ms after stimulus onset. While the electrodogram outputs suggest that cochlear implant provides some access to spectral cues that distinguish the two consonants to children with BiCIs in this study, it is likely that the spectral cues available through the processors were too coarse for the saliency they needed to achieve the same level of performance as NH children. Currently, most cochlear implants use monopolar stimulation strategy which has wide spread of excitation along the cochlea that results in reduced spectral resolution. New stimulation strategies using tripolar and partial tripolar stimulation with electrical current focusing were previously shown to improve spectral resolution among adults with cochlear implants by reducing spread of excitation (21). As finer spectral resolution may improve the saliency of fine-grained spectral cues in the consonant contrasts used in this study, one future direction may be to understand the benefit of focused stimulation strategies in these very young children with cochlear implants.

For the voicing contrast, the VOT cue was preserved in a fairly intact manner, with a 50 ms delay in the voicing /b/-/p/ contrast as seen in the simulated electrodograms. Previous studies showed that older children with cochlear implants between 5 and 7 years had similar accuracy as NH children in discriminating VOT delays less than 50 ms (10,11). While implant processors were able to preserve the VOT cue, the specific VOT cue of 50 ms delay might not have been salient enough for 2-year old children with BiCIs to arrive at the same level of discrimination accuracy as their peers with NH. For younger children with BiCIs (2 yr old in this study), the VOT cue of 50 ms seemed to drive performance above chance level but was not strong enough to provide the same discrimination accuracy as their NH peers. As expected, with a shorter VOT delay of 20 ms in the modified /b/-/p/ contrast, in which the VOT cue became less salient, both groups of children performed more poorly; yet children with BiCIs still performed above chance levels with this cue.

We observed substantial individual variability in performance of children with BiCIs in consonant contrast discrimination, with some children performing close to chance and others nearly perfect. Even with a small sample size, there was a trend toward improved discrimination accuracy among the older children, specifically in the place /p/-/k/ contrast. These results suggest that some children with BiCIs might eventually catch up with their NH peers in discriminating consonant contrasts.

Acknowledgments: The authors thank Shelly Godar for audiological assistance, Alyssa Lamers, Erica Ehlers, and Kayla Kristensen for assistance with data collection, and Matthew Winn for assistance with stimulus generation. Electrodogram simulators were provided by Advanced Bionics and Cochlear. They are grateful for the families who traveled to Madison, WI and those from the local community to participate in this study.

REFERENCES

1. Boothroyd A. Auditory capacity of hearing-impaired children using hearing aids and cochlear implants: issues of efficacy and assessment. *Scand Audiol Suppl* 1997;26:17–25.
2. O'Donoghue GM, Nikolopoulos TP, Archbold SM, et al. Cochlear implants in young children: the relationship between speech perception and speech intelligibility. *Ear Hear* 1999;20:419–25.

3. Blamey PJ, Sarant JZ, Paatsch LE, et al. Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing. *J Speech Lang Hear Res* 2001;44:264.
4. Niparko JK, Tobey EA, Thal DJ, et al. Spoken language development in children following cochlear implantation. *JAMA* 2010;303:1498–506.
5. Dunn CC, Walker EA, Oleson J, et al. Longitudinal speech perception and language performance in pediatric cochlear implant users: the effect of age at implantation. *Ear Hear* 2014;35:148–60.
6. Moffitt AR. Consonant cue perception by twenty- to twenty-four-week-old infants. *Child Dev* 1971;42:717–31.
7. Eilers RE, Wilson WR, Moore JM. Developmental changes in speech discrimination in infants. *J Speech Hear Res* 1977;20:766–80.
8. Eilers RE, Morse PA, Gavin WJ, Oller DK. Discrimination of voice onset time in infancy. *J Acoust Soc Am* 1981;70:955–65.
9. Aslin RN, Pisoni DB, Hennessy BL, Perey AJ. Discrimination of voice onset time by human infants: new findings and implications for the effects of early experience. *Child Dev* 1981;52:1135–45.
10. Giezen MR, Escudero P, Baker A. Use of acoustic cues by children with cochlear implants. *J Speech Lang Hear Res* 2010;53:1440–57.
11. Caldwell A, Nittrouer S. Speech perception in noise by children with cochlear implants. *J Speech Lang Hear Res* 2013;56:13–30.
12. Loizou PC, Hu Y, Litovsky R, et al. Speech recognition by bilateral cochlear implant users in a cocktail-party setting. *J Acoust Soc Am* 2009;125:372–83.
13. Raggio MW. Neuronal responses in cat primary auditory cortex to electrical cochlear stimulation: IV. Activation pattern for sinusoidal stimulation. *J Neurophysiol* 2003;89:3190–204.
14. Middlebrooks JC, Bierer JA, Snyder RL. Cochlear implants: the view from the brain. *Curr Opin Neurobiol* 2005;15:488–93.
15. Perris EE, Clifton RK. Reaching in the dark toward sound as a measure of auditory localization in infants. *Infant Behav Dev* 1988;11:473–91.
16. Clifton RK, Rochat P, Litovsky RY, Perris EE. Object representation guides infants' reaching in the dark. *J Exp Psychol Hum Percept Perform* 1991;17:323–9.
17. Litovsky RY, Ehlers E, Hess C, Harris S. Reaching for sound measures: an ecologically valid estimate of spatial hearing in 2- to 3-year-old children with bilateral cochlear implants. *Otol Neurotol* 2013;34:429–35.
18. Coady JA, Kluender KR, Evans JL. Categorical perception of speech by children with specific language impairments. *J Speech Lang Hear Res* 2005;48:944–59.
19. Boersma P. Praat, a system for doing phonetics by computer. *Glot Int* 2001;5:341–7.
20. Studebaker GA. A ''rationalized'' arcsine transform. *J Speech Hear Res* 1985;28:455–62.
21. Berenstein CK, Mens LHM, Mulder JJS, Vanpoucke FJ. Current steering and current focusing in cochlear implants: comparison of monopolar, tripolar, and virtual channel electrode configurations. *Ear Hear* 2008;29:250–60.