Vowel discrimination by hearing infants as a function of number of spectral channels[a]

Andrea D. Warner-Czyz[b]
Department of Communication Sciences and Disorders, The University of Texas at Dallas,
Callier Advanced Hearing Research Center, 1966 Inwood Road, Dallas, Texas 75235

Derek M. Houston
Department of Otolaryngology, Head and Neck Surgery, Indiana University School of Medicine,
699 Riley Hospital Drive/RR044, Indianapolis, Indiana 46202

Linda S. Hynan
Departments of Clinical Sciences and Psychiatry, The University of Texas Southwestern Medical Center,
5323 Harry Hines Boulevard, Dallas, Texas 75390

Reduced spectral resolution negatively impacts speech perception, particularly perception of vowels and consonant place. This study assessed impact of number of spectral channels on vowel discrimination by 6-month-old infants with normal hearing by comparing three listening conditions: Unprocessed speech, 32 channels, and 16 channels. Auditory stimuli (/ti/ and /ta/) were spectrally reduced using a noiseband vocoder and presented to infants with normal hearing via visual habituation. Results supported a significant effect of number of channels on vowel discrimination by 6-month-old infants. No differences emerged between unprocessed and 32-channel conditions in which infants looked longer during novel stimulus trials (i.e., discrimination). The 16-channel condition yielded a significantly different pattern: Infants demonstrated no significant difference in looking time to familiar vs novel stimulus trials, suggesting infants cannot discriminate /ti/ and /ta/ with only 16 channels. Results support effects of spectral resolution on vowel discrimination. Relative to published reports, young infants need more spectral detail than older children and adults to perceive spectrally degraded speech. Results have implications for development of perception by infants with hearing loss who receive auditory prostheses. © 2014 Acoustical Society of America.
[http://dx.doi.org/10.1121/1.4870700]

I. INTRODUCTION

Speech perception is a complex cognitive task that requires detailed spectral and temporal information to recognize the acoustic patterns of speech. The brain's ability to identify patterns in a signal relies on basic perceptual properties like speech discrimination. Signal degradations such as reduced spectral detail received via a cochlear implant (CI) may negatively impact speech discrimination, creating difficulty in acquiring perceptual skills in individuals with hearing loss who receive CIs. Reduced spectral resolution may particularly impact infants, who—having less experience with language input and less time to develop top-down processing skills relative to adults—may need more complete sensory input to perceive stimuli properties that drive bottom-up processing. That is, in the absence of top-down processing, infants may rely more heavily on bottom-up processing until they can recognize patterns in the signal.

A child's ability to resolve an impoverished signal received via CI may contribute to characteristic variability in perception and language outcomes (Geers *et al.*, 2003; Geers *et al.*, 2008). CIs restore the sensation of hearing to persons with hearing loss with spectral degradation determined by the number of stimulated nodes on the electrode array implanted in the cochlea. Reduced spectral resolution affects frequency discrimination, which underlies location of spectral peaks (e.g., first and second formants), cues important for identification of vowels and consonant place (Rosen, 1992; Stevens, 1980, 1983). Difficulty perceiving consonants and vowels within syllables may have cascading effects from word segmentation based on a signal's phonotactic regularities to attainment of complex language skills (word recognition, syntax, morphology, reading) (Geers and Hayes, 2011; Jusczyk, 1985; Mattys *et al.*, 1999; Niparko *et al.*, 2010; Saffran *et al.*, 1996; Svirsky *et al.*, 2002).

One way to isolate the impact of reduced spectral resolution from effects of auditory deprivation and cognitive issues on speech perception is to present a CI-simulated signal to listeners with normal hearing (NH). Spectral reduction techniques divide an acoustic signal into discrete frequency channels using multiple bandpass filters while preserving temporal envelope modulations of speech (Shannon *et al.*, 1995). A positive relationship exists between number of channels and perception in CI simulations in adults: Fewer

---

channels correspond to lower accuracy (e.g., Baskent, 2006; Dorman *et al.*, 1997; Eisenberg *et al.*, 2000; Shannon *et al.*, 1995). The number of spectral channels required for high performance accuracy differs by stimulus type. NH adults achieve excellent discrimination (>90% correct) of consonant voicing and manner—which primarily rely on temporal cues—with two channels (Shannon *et al.*, 1995). NH adults achieve excellent vowel recognition, which relies more on spectral cues, with three to four channels, given extensive training and practice with spectrally reduced stimuli (Eisenberg *et al.*, 2000; Shannon *et al.*, 1995). Given little to no practice with spectrally reduced vowels, however, adults need 5–12 channels to score greater than 90% correct (Dorman *et al.*, 1997; Friesen *et al.*, 2001; Xu *et al.*, 2005).

A. Perception of vocoded speech by children (3–13 yr of age)

An inverse relationship exists between age and number of channels for excellent scores (defined as >90% correct) on an age-appropriate speech perception task in 3- to 13-yr-old children (Bertoncini *et al.*, 2009; Dorman *et al.*, 2000; Eisenberg *et al.*, 2000; Nittrouer and Lowenstein, 2010; Nittrouer *et al.*, 2009). Pre-teens (10–13 yr) have adult-like performance, requiring six noiseband vocoded channels on a composite of syllable contrasts (two vowels, six consonants) and four noiseband channels for vowel contrasts (/ti/-/ta/ and /ti/-/tu/) (Eisenberg *et al.*, 2000). Younger children, however, needed more spectral detail to achieve similar performance levels. Scores for NH 5- to 7-yr-old children did not exceed 80% correct for a composite of syllable contrasts with eight channels. The younger children achieved a mean score of 90% correct on vowel contrasts with six noiseband channels—two more channels than required by pre-teens and adults to achieve excellent discrimination scores (Eisenberg *et al.*, 2000).

The relationship between age and perception of vocoded speech persists in younger NH children. Preschool-aged children (3–5 yr) do not attain adult-like perception of vocoded speech (Dorman *et al.*, 2000; Nittrouer and Lowenstein, 2010). For example, NH preschoolers need more spectral cues than adults (12–20 vs 10 channels) to score >90% correct on multisyllabic words reduced via a sinewave vocoder (Dorman *et al.*, 2000; Eisenberg *et al.*, 2000). Comparison across studies reveals a consistent negative association between age and number of channels required to achieve excellent speech perception scores.

B. Perception of vocoded speech by infants (<1 yr of age)

Infants with NH discriminate many phonetic contrasts by age 6 months (e.g., Eilers *et al.*, 1977; Eimas *et al.*, 1971; Jusczyk *et al.*, 1978; Kuhl, 1979). Only one study thus far explores the effect of spectral reduction on infant perception (Bertoncini *et al.*, 2011). Bertoncini *et al.* (2011) tested discrimination of consonant voicing (/aba/ vs /apa/) with 16 sinewave-vocoded channels in 20 6-month-old NH infants using a head-turn preference method that contrasts alternating and repeating stimuli. Seventy percent of infants looked longer during alternation vs repetition series, implying

sensitivity to differences in envelope cues. The authors concluded 6-month-old infants do not need fine spectral or temporal information for speech discrimination.

Spectral reduction should minimally affect perception of consonant voicing, which primarily relies on temporal cues (i.e., voice onset time) with little impact from spectral cues. These results converge with Eisenberg *et al.* (2000), who found that pediatric and adult listeners accurately discriminated an easy voicing contrast (/da/-/ta/) with fewer than eight channels. To date, no studies have explored discrimination of any other contrasts in young infants.

C. The present experiment

Very little is known about infants' ability to perceive spectrally degraded speech. Gaining this knowledge is crucial because infants comprise 20% of CI recipients as a result of earlier identification of hearing loss and audiological intervention. Clinical and research teams routinely implant children younger than the Food and Drug Administration's approved age of 12 months. To better serve this pediatric population, we need to understand the role of spectral resolution in perception to inform the minimum criterion of channels provided to CI in infants.

An early step to answering this question is to determine how many channels infants with NH need to discriminate a degraded and unprocessed signal—not only for contrasts that rely primarily on temporal cues (i.e., consonant voicing) but also for contrasts that depend primarily on spectral cues (i.e., vowels). If number of channels alone determines ability to discriminate CI-simulated speech, then infants should discriminate vowel and consonant voicing contrasts with 16 channels equally. Alternatively, if stimulus contrast impacts perception of vocoded speech, then infants should discriminate vowels differently than consonant voicing. This could mean greater difficulty discriminating vowel vs consonant voicing contrasts, requiring more channels to discriminate vowels as demonstrated by CI simulations with NH adults (Shannon *et al.*, 1995); or infants listening to CI simulations could perform more similarly to adult CI users, who exhibit greater difficulty perceiving consonant voicing—especially for fricatives—compared to some vowel contrasts (Munson *et al.*, 2003). The present study explores how much signal degradation, like that via CI, an immature brain can process by asking, Does the level of spectral resolution (i.e., 16 spectral channels, 32 spectral channels, or unprocessed signal) impact discrimination of spectrally distinct syllables (/ti/ vs /ta/) in NH 6-months-old infants?

II. METHODS

A. Participants

We tested 6-month-old infants to match the age of infants in Bertoncini *et al.* (2011). Participants included 52 5- to 7-month-old infants (34 male, 18 female) recruited from the Dallas metropolitan area (Table I). Infant ages ranged from 5 months, 0 days to 7 months, 30 days ($M = 190$ days, SD = 26.0). All infants met the following inclusion criteria: Normal peripheral auditory function assessed via

TABLE I. Demographic variables of 5- to 7-month-old infant participants.

Variable	N	Percentage
Language environment		
English-learning	39	75.00
Spanish-learning	13	25.00
Ethnicity		
Hispanic or Latino	31	59.62
Not Hispanic or Latino	17	32.69
Unknown or not reported	4	7.69
Race		
Asian or Pacific Islander	2	3.85
Black or African-American	5	9.62
White or Caucasian	17	32.69
More than one race reported	2	3.85
Unknown or not reported	26	50.00

distortion-product otoacoustic emissions screening; normal middle ear status confirmed via 1000 Hz tympanometry; product of full-term pregnancy (≥37 weeks gestation); reared in an English- or Spanish-learning home environment; habituation to a familiar stimulus (see Sec. II C). Infants with diagnosed or suspected cognitive, visual, or developmental delay and infants with a history of otitis media were excluded. Data from an additional 32 infants were removed from analysis due to fussiness ($n = 7$), inattention ($n = 4$), inability to habituate ($n = 8$), or experimental error ($n = 13$). Failure to habituate occurred more frequently with spectrally degraded speech ($n = 6$ and 2 in the 16- and 32-channel conditions, respectively). Each family received $20 for participation in the 1-h visit. This project was approved by the Institutional Review Board at The University of Texas at Dallas (Protocol 09-08).

Participants were assigned randomly to a listening condition (16 channels, 32 channels, unprocessed) (see Sec. II B 2). Mean age (with standard deviations in parentheses) was similar across the three groups: 192.00 days (SD = 30.36), 190.00 days (SD = 20.77), and 185.18 days (SD = 28.62) for the 16 channel, 32 channel, and unprocessed conditions, respectively. Most infants across conditions (75%) were being reared in an English-learning environment (85% in the 16-channel condition; 62% in the 32-channel condition; and 82% in the unprocessed condition). Maternal education level ranged from completion of 7th grade to earning a college degree. One-third of the infants' mothers obtained at least a Bachelor's degree (35%) and nearly one-third graduated from high school (31%). Annual income ranged from less than $15 000 to greater than $130 000, with nearly equal distribution of families earning less than $30 000 (31%), between $30 000 and $75 000 (26%), and greater than $75 000 (29%) per year (13% declined to answer).

B. Data collection

1. Apparatus

We conducted testing in a double-walled Industrial Acoustics Company soundproof booth. Infants sat on a caregiver's lap 4 ft from a 46-in. Samsung 460 CX LCD monitor.

Visual stimuli were centered on the monitor at the infant's eye level. We presented auditory stimuli via a Logitech Z-2300 multimedia speaker system with two speakers mounted 22 in. apart atop of the monitor. The experimenter observed infants from a separate room via closed-circuit input from two Canon VC-C50i communication cameras using Noldus MPEG recorder software on a Dell Precision T3400 computer. One camera focused on the infant's face; the other on the screen. The tester controlled the experiment using HABIT X 1.0 software (University of Texas Children's Research Laboratory, Austin, TX) on a Macintosh MacPro G6 computer.

2. Stimuli

Consonant-vowel stimuli were selected to represent the natural phonotactic syllable structure of consonants and vowels in speech versus isolated vowels. Stimuli were constructed of highly contrastive phonemic units, /ti/ and /ta/, to highlight spectral differences among vowels that constitute the articulatory and acoustic extremes of the vowel space. Five unique tokens of each syllable were selected from 10 audio recordings of a single female talker instructed to produce speech contrasts as if speaking to an infant. This multiple natural tokens approach reflects our intent to assess discrimination of contrasts in which minor acoustic variations should be immaterial (Kuhl, 1979; Werker et al., 1981). Recordings were digitized at a sampling rate of 44.1 kHz with 16-bit resolution using an Edirol R-09HR high-resolution wave/MP3 recorder. All stimuli tokens were equated for duration (500 ms), total rms power, and intensity [65 ± 5 dB sound pressure level (SPL)] using ADOBE AUDITION 3.0. Stimuli were separated by 500 ms inter-stimulus interval.

The stimuli described in the preceding text comprised the "unprocessed" listening condition. The same auditory stimuli were spectrally reduced by implementing an n-channel noiseband vocoder using MATLAB 7.8 (Dorman et al., 1997; Shannon et al., 1995). A noiseband vocoder digitally creates spectral channels using filtered noisebands while maintaining temporal information (i.e., speech amplitude envelope) within each spectral band channel. We selected a noiseband vocoder to mirror previously published studies evaluating perception of spectrally reduced vowels, which predominantly use noiseband vocoders (e.g., Baskent, 2006; Eisenberg et al., 2000; Friesen et al., 2001; Fu et al., 1998; Shannon et al., 1995; Xu and Zheng, 2007) versus sinewave (Dorman and Loizou, 1998) or both types of vocoders (Dorman et al., 1997). Spectral reduction was accomplished by subjecting the unprocessed digitized acoustic signal to a pre-emphasis filter followed by a set of bandpass filters. The pre-emphasis filter had a cut-off frequency of 1200 Hz with a 6 dB/octave roll-off below 1200 Hz. The filtered signal was input into a series of either 16 or 32 bandpass filters implemented using sixth order Butterworth filters. To detect the speech amplitude envelope, output from each bandpass filter then was processed via half-wave rectification followed by low-pass filtering using second order Butterworth filters with a 160 Hz cut-off frequency. The extracted envelopes from each channel modulated white noise that was band-limited to the same bandpass filters as the original filters. Channels

J. Acoust. Soc. Am., Vol. 135, No. 5, May 2014

Warner-Czyz et al.: Infant perception of spectrally reduced cue   3019

were distributed by semi-logarithmic (mel) spacing for both the 16- and 32-channel conditions due to systematic ability to compute filter bandwidths and its similarity to filter spacing used in CI devices (Loizou, 1998). Table II shows frequency allocation tables with the center frequency for each channel in the two spectrally reduced listening conditions.

Three visual stimuli were used during testing. First, a silent dynamic video of an infant laughing was used to orient the infant to the monitor before each trial (*attention getter*). Second, infants attended to a 4 × 5 black and white checkerboard while listening to stimuli during habituation and test trials (*simple visual stimulus*). Third, a computer graphic animation paired with a sequence of brief pure-tone stimuli gauged the infant's general attention and arousal level before and after the experiment (*attention measure*) (see next section for details).

C. Procedure: Visual habituation

We used visual habituation, a well-established methodology to assess discrimination in children younger than 6 months. All infants were assigned randomly to one listening condition as a standard procedure to decrease sampling bias. Testing consisted of two phases: A habituation and a test phase (Houston *et al.*, 2007). In the *habituation phase*, the experimenter presented up to 16 trials that paired repetitions of one stimulus (e.g., /ti ti ti…/) with a simple visual stimulus (checkerboard). The habituation phase continued until the infant's mean looking time during three consecutive trials decreased to 50% of mean looking time during the three longest trials, a standard decrement in infant perception studies (Cohen *et al.*, 1975). If infants did not meet this criterion, we considered the habituation unsuccessful and excluded data from analysis. We used infant-controlled habituation criteria to equate subjects for the extent to which the familiar stimulus was encoded, increasing the chance infants would look longer during novel stimulus trials after the habituation phase (Cohen, 1972; Horowitz *et al.*, 1972). The test phase began after an infant habituated. In the *test phase*, the experimenter presented 12 trials in an oddball paradigm with 9 trials of the familiar stimulus (e.g., /ti ti ti…/) and 3 trials of the novel stimulus (e.g., /ta ta ta…/) (Houston *et al.*, 2007). Test trials were divided into three blocks of three familiar trials and one novel trial. Stimulus order was randomized within blocks and counterbalanced for presentation order (i.e., /ti/ as the familiar stimulus with /ta/ as the novel stimulus and vice versa) with the caveat that two novel trials never occurred consecutively.

Infant looking times were assessed during the experiment using HABIT X 1.0 software (University of Texas Children's Research Laboratory, Austin, TX). To eliminate tester bias, the tester wore headphones and was blind to stimulus presentation order, simply pressing a computer key when the infant's eyes oriented toward the monitor. To limit caregiver bias, the caregiver listened to music and was instructed to refrain from modifying infant attention by pointing to the monitor, turning the child's head toward the monitor, etc. The attention getter was used to orient the infant to the monitor before each trial onset. The trial began once the infant oriented to the monitor and continued until the infant looked away from the monitor for longer than 1 s or until the maximum trial length of 30 s was reached. Infant looking time for each trial was calculated as total amount of time fixated on the visual stimulus during the trial.

We quantified general attention and arousal level of the infants using the previously described attention measure, which was completely unrelated to the experimental stimuli, to assure decreased looking time at the end of the experiment related to response to the stimulus and not decrease in overall attention or general fatigue (Cohen and Amsel, 1998).

D. Data analysis

Once testing was complete, we calculated the *difference in mean looking time* between novel and familiar stimulus trials. Due to infants' proclivity to pay attention to novel stimuli, we expect infants to look longer during novel stimulus trials if they can discriminate the two vowels. A preference for a novel stimulus occurs when an infant has encoded a robust enough memory representation of the familiar stimulus during the habituation phase to habituate to it and thus attend more to the novel stimulus during the test phase (Houston-Price and Nakai, 2004; Roder *et al.*, 2000; Wetherford and Cohen, 1973). No difference in mean looking time was interpreted as an inability to discriminate the speech contrast. Ten percent of the videos were reviewed by an independent trained laboratory associate. Calculation of Cronbach's alpha for the reliability statistic revealed excellent agreement between the two coders (Cronbach's $\alpha = 0.98$).

III. RESULTS

This experiment examined the relationship between spectral resolution and speech discrimination in 6-month-old NH infants. A two-way analysis of variance (ANOVA) with two between-subjects factor (*listening condition*: Unprocessed speech, 16 channels, or 32 channels; *presentation order*: /ti/-/ta/ vs /ta/-/ti/) evaluated the effect of listening condition on difference in mean looking time with an alpha value of 0.05 using SAS V9.3 (SAS Institute, Inc., Cary, NC). The ANOVA was run through a general linear model to adjust for unequal observations per cell.

TABLE II. Center channel frequencies (in Hz) for spectrally reduced signals with 16 and 32 channels.

No. of channels	Center frequencies (Hz)															
16 channels	216	343	486	647	828	1031	1260	1518	1808	2134	2501	2914	3378	3901	4489	5150
32 channels	132	190	252	318	387	462	541	625	714	809	909	1016	1130	1251	1379	1575
	1660	1814	1978	2151	2336	2532	2741	2962	3198	3448	3714	3996	4297	4616	4954	5315

Listening condition significantly impacted difference in mean looking time, $F(2, 46) = 4.65$, $p = 0.01$. Six-month-old infants had longer looking times during novel stimulus trials with unprocessed speech ($M = 1.46$ s, SD = 2.59) and 32 channels ($M = 1.34$ s, SD = 2.49) but longer mean looking times during familiar stimulus trials with 16 channels [$M = -0.64$ s, SD = 1.89 (Fig. 1)]. Planned Bonferroni multiple comparisons indicate that the 16-channel condition differs significantly from the 32-channel ($p = 0.02$) and unprocessed listening conditions ($p = 0.04$), but that the 32-channel and unprocessed listening conditions do not differ significantly from each other ($p > 0.99$). Further analysis reveals no significant effects of presentation order and no significant interaction between listening condition and presentation order.

To determine if infants discriminated the speech contrast in each condition, we ran a two-way mixed ANOVA with one between-subjects variable (i.e., listening condition) and one within-subjects variable (i.e., mean looking time during familiar vs novel stimulus trials). A subsequent simple effects test compared mean looking time to familiar vs novel stimulus trials within each listening conditions. Results confirm infants looked significantly longer during the novel stimulus trials—evidence of /ti/-/ta/ discrimination – for the unprocessed, $F(1, 49) = 4.44$, $p = 0.040$ and 32-channel conditions, $F(1, 49) = 7.18$, $p = 0.010$. Infants did not show a significant difference in looking time for the 16-channel condition [$F(1, 49) = 1.55$, $p = 0.219$], suggesting that infants did not discriminate /ti/-/ta/ with 16 channels,.

We conducted an additional ANOVA to ensure differences by listening condition did not relate to differences in other factors such as number of trials to habituation, attention and arousal. Mean looking time during the pre-test attention measure ($M = 24.70$, SD = 7.85) exceeded that of the post-test attention measure ($M = 17.49$, SD = 3.94) across conditions.

However, no significant differences between listening conditions emerged for any of these factors.

IV. DISCUSSION

This study investigated how much spectral resolution 6-month-old infants with NH need to discriminate vowels (/ti/-/ta/) by comparing three listening conditions: Unprocessed speech, 32 channels, and 16 channels, in decreasing order of spectral resolution. Infants looked significantly longer during novel stimuli trials in the unprocessed and 32-channel conditions, indicating they could discriminate the vowel contrast in these two listening conditions. In contrast, infants demonstrated no difference in looking time between the familiar and novel stimuli in the 16-channel condition, suggesting infants could not discriminate the vowel contrast in the most impoverished listening condition. Thus our results suggested infants discriminate /ti/-/ta/ with unprocessed and 32-channel speech but not with a 16-channel signal.

Common variables contributing to differences in looking patterns in visual habituation paradigms include age, familiarization time, and task difficulty (e.g. Hunter and Ames, 1988). Infants across listening conditions were essentially the same age so it is improbable age contributed to group discrepancies. Infants in the 16-channel condition may have discriminated /ti/-/ta/ had we allowed a longer familiarization time (i.e., >16 habituation trials) (Hunter and Ames, 1988). However, we used infant-controlled habituation criteria, which reduced the chances that some infants were not fully habituated (Cohen, 1972; Horowitz et al., 1972). Also, we excluded all infants who did not reach the habituation criteria within 16 trials.

The most likely explanation for the switch from novelty preferences for the unprocessed and 32-channel conditions
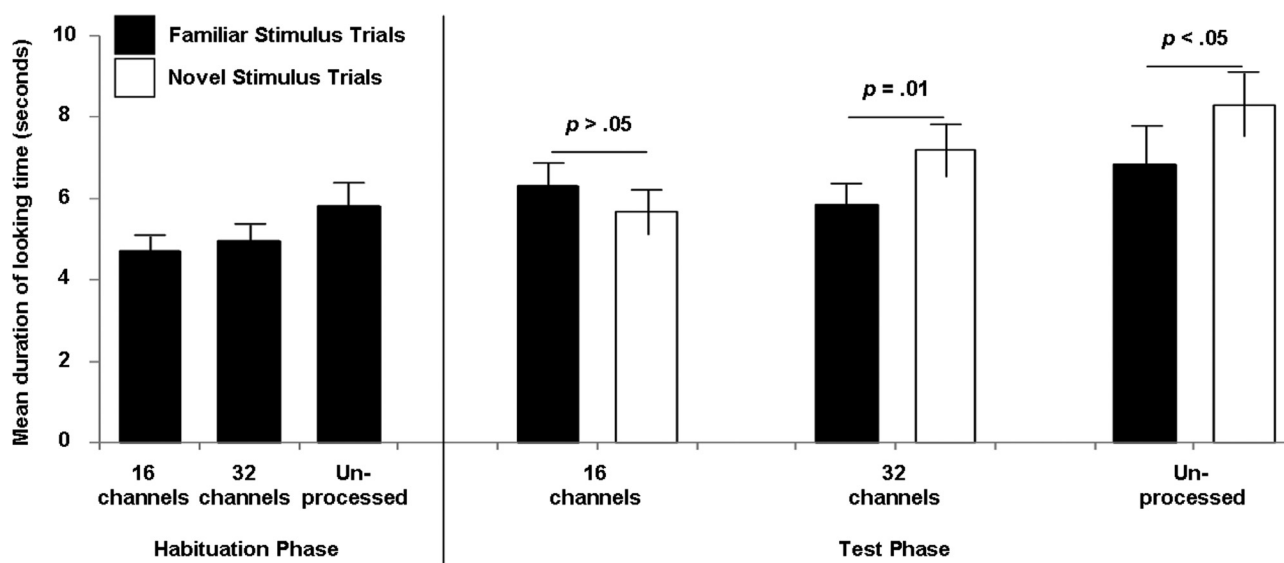

FIG. 1. Mean duration of looking time during visual habituation testing listening to a spectrally distinct vowel contrast (/ti/ vs /ta/) across three listening conditions: Unprocessed speech, 32 spectral channels, and 16 spectral channels, in decreasing order of spectral resolution. The left side of the figure displays mean looking times during the final three trials of the habituation phase, which only includes the familiar stimulus, for each listening condition. The right side of the figure shows mean looking times during familiar stimulus trials (filled columns) and novel stimulus trials (open columns). A significant difference in mean looking time during familiar versus novel stimulus trials within a given listening condition implies discrimination of the contrast. No significant difference in mean looking time during familiar versus novel stimulus trials indicates no discrimination of the contrast. The bars represent 1 standard error of the mean.

to no preference for the 16-channel condition is an increased difficulty encoding and discriminating the vowels when given less spectral information. Fewer spectral bands with 16 vs 32 channels means each channel included a broader frequency range. This increases the chance that formants will activate adjacent rather than separate frequency bands, potentially creating similar coding patterns for /ti/ and /ta/. Also ability to habituate varied by condition, with 75% of non-habituators in the most difficult, 16-channel condition and the remaining 25% in the 32-channel group, suggesting that the 16-channel stimuli were more difficult to encode than the 32-channel stimuli. This possibility is bolstered further by pilot data with eight-channel stimuli. None of the four infants met the habituation criteria with eight channels, consistent with the interpretation that infants require more information (i.e., longer habituation phase, more channels) to encode an increasingly degraded signal.

Our findings that 6-month-old infants with NH cannot discriminate vowels (/ti/-/ta/) with 16 channels oppose results from Bertoncini et al. (2011), who reported 6-month-old infants with NH can discriminate consonant voicing (/aba/ vs /apa/) with 16 channels. Several factors could explain these divergent results. The studies used different methodologies to test infant discrimination: Visual habituation (this study) and a head-turn procedure (Bertoncini et al., 2011). Both methods rely on an infant's intrinsic interest in novel stimuli and are appropriate to assess speech discrimination in 6-month-old infants (Houston-Price and Nakai, 2004; Polka and Werker, 1994; Werker et al., 1998). Therefore differences likely reflect true differences based on factors other than methodology.

Stimulus contrast represents another difference between the two infant studies. Bertoncini et al. (2011) presented a disyllabic consonant voicing contrast (/apa/-/aba/), which relies on temporal differences in the acoustic signal (i.e., voice onset time) with little input from the spectral domain. We used a monosyllabic vowel contrast (/ti/-/ta/), which depends more on spectral information (Dorman et al., 1997; Rosen, 1992). These basic perceptual characteristics set the expectation of vowel perception as a more difficult task when degrading spectral resolution. The need for more spectral channels to discriminate vowels versus consonant voicing, shown here in infants, mirrors results in CI simulation studies involving older children and adults with NH (Baskent, 2006; Dorman et al., 1997; Eisenberg et al., 2000; Friesen et al., 2001; Schvartz et al., 2008; Shannon et al., 1995; Xu et al., 2005).

A. CI simulations as a model for communication performance in CI recipients

Application of CI simulations as a model for speech perception performance in CI users requires more consideration. Studies generally report similar perception scores for CI recipients and NH listeners tested with spectrally reduced stimuli, although differences exist (Dorman and Loizou, 1997; Eisenberg et al., 2002; Fu et al., 1998). For example, adults with NH require more channels to achieve excellent perception of vowels than of consonant voicing with CI

simulations, but adults using CI exhibit a different response pattern. Munson et al. (2003) found adult CI users had more misperceptions of consonants than vowels (e.g., 44 misperceptions of the /b/-/p/ contrast and 0 misperceptions of the /a/-/i/ contrast). These participants, however, were post-lingually deafened adults with a mean of 4 yr of experience with the CI—a different population than young infants listening to a spectrally degraded signal for the first time.

Pediatric CI users also may acquire speech perception skills differently than expected from CI simulations. Uhler et al. (2011) showed that infants who receive CI between 12 and 16 months of age can discriminate three of five speech contrasts (/a/-/i/, /a/-/u/, /u/-/i/, /sa/-/ma/, /pa/-/ka/) with 3 months of listening with the CI, although the combination of contrasts varied by child. Blamey and colleagues (Blamey et al., 2001) reported similar findings in their work on early phoneme acquisition: More than half of the nine children implanted by age 5 yr produced /a/ pre-implant and /i/ by 6 months post-implantation. Thus pediatric CI users' ability to discriminate and produce consonant and vowel contrasts emerges before 6 months of device use, presumably due to consistent access to a spectrally degraded signal that the brain learns to interpret as speech.

Performance disparities call into question the representativeness of CI simulations. Neither sinewave nor noiseband vocoders imitate all aspects of a CI. Sinewave vocoders contain harmonic-like components centered within each channel, which highlights cues for voicing and intonation (Souza and Rosen, 2009). Noiseband vocoders elicit auditory nerve responses similar to natural speech (Loebach and Wickesburg, 2006). Neither spectral reduction method perfectly mimics a CI signal, but each vocoder type can approximate outcomes for CI users.

Although models of CI simulation potentially underestimate how well an infant or toddler with CI may perform, they provide a baseline for comparison based not only on chronologic age but also listening age. Results from this study suggest that 6-month-old infants may need more than 16 spectral channels to accomplish an age-appropriate speech perception task such as syllable discrimination with a CI. These findings should be taken into consideration when fitting young infants with contemporary CI devices, which include between 12 and 22 electrode contacts. Likewise, the first 6 months post-implant may require additional spectral information to allow the infant to develop the bottom-up processing skills necessary to learn language.

B. Future directions

The trajectory by which infants with CI acquire speech discrimination skills may differ from infants with NH. Typically developing infants with NH discriminate most speech sounds at birth and narrow their focus to sounds of their native language between 6 and 12 months (Jusczyk et al., 1994; Werker and Tees, 1984). In contrast, infants with CI begin with little to no perceptual abilities and gain discrimination capacities with CI experience. This suggests that the development of speech perception skills may not

only be delayed in CI users but follow a very different trajectory that could influence how they acquire language.

This study presents a way to evaluate impact of reduced spectral resolution on speech discrimination abilities. Future studies using CI simulations should consider impact of both task difficulty (e.g., number of channels, stimuli type) and age to document emergence of perception with a degraded signal over the course of childhood. Additionally, future research should expand to include infants with hearing loss using hearing aids and CI to determine similarities and differences in performance relative to infants and children with NH and to determine accuracy of CI simulations to actual pediatric CI recipients.

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

Bertoncini, J., Nazzi, T., Cabrera, L., and Lorenzi, C. (**2011**). "Six-month-old infants discriminate voicing on the basis of temporal envelope cues (L)," J. Acoust. Soc. Am. **129**, 2761–2764.

Bertoncini, J., Serniclaes, W., and Lorenzi, C. (**2009**). "Discrimination of speech sounds based upon temporal envelope versus fine structure cues in 5- to 7-year-old children," J. Speech Lang. Hear. Res. **52**, 682–695.

Blamey, P. J., Barry, J. G., and Jacq, P. (**2001**). "Phonetic inventory development in young cochlear implant users 6 years postoperation," J. Speech Lang. Hear. Res. **44**, 73–79.

Cohen, L. B. (**1972**). "Attention-getting and attention-holding processes of infant visual preferences," Child Dev. **43**, 869–879.

Cohen, L. B., and Amsel, G. (**1998**). "Precursors to infants' perception of the causality of a simple event," Infant Behav. Dev. **21**, 713–731.

Cohen, L. B., DeLoache, J. S., and Rissman, M. W. (**1975**). "The effect of stimulus complexity on infant visual attention and habituation," Child Dev. **46**, 611–617.

Dorman, M. F., and Loizou, P. C. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants," Am. J. Otol. **18**, S113–S114.

Dorman, M. F., and Loizou, P. C. (**1998**). "The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels," Ear Hear. **19**, 162–166.

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

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**, 2403–2411.

Eilers, R. E., Wilson, W. R., and Moore, J. M. (**1977**). "Developmental changes in speech discrimination in infants," J. Speech Hear. Res. **20**, 766–780.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (**1971**). "Speech perception in infants," Science **171**, 303–306.

Eisenberg, L. S., Martinez, A. S., Holowecky, S. R., and Pogorelsky, S. (**2002**). "Recognition of lexically controlled words and sentences by children with normal hearing and children with cochlear implants," Ear Hear. **23**, 450–462.

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

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

Fu, Q. J., Shannon, R. V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.

Geers, A., Brenner, C., and Davidson, L. (**2003**). "Factors associated with development of speech perception skills in children implanted by age five," Ear Hear. **24**, 24S–35S.

Geers, A., Tobey, E., Moog, J., and Brenner, C. (**2008**). "Long-term outcomes of cochlear implantation in the preschool years: From elementary grades to high school," Int. J. Audiol. **47** Suppl. 2, S21–S30.

Geers, A. E., and Hayes, H. (**2011**). "Reading, writing, and phonological processing skills of adolescents with 10 or more years of cochlear implant experience," Ear Hear. **32**, 49S–59S.

Horowitz, F. D., Paden, L., Bhana, K., and Self, P. (**1972**). "An infant control method for studying infant visual fixations," Dev. Psychol. **7**, 90

Houston, D. M., Horn, D. L., Qi, R., Ting, J., and Gao, S. (**2007**). "Assessing speech discrimination in individual infants," Infancy **12**, 119–145.

Houston-Price, C., and Nakai, S. (**2004**). "Distinguishing novelty and familiarity effects in infant preference procedures," Infant Child Dev. **13**, 341–348.

Hunter, M., and Ames, E. (**1988**). "A multifactor model of infant preferences for novel and familiar stimuli," in Advances in Infancy Research, edited by C. Rovee-Collier and L. P. Lipsitt (Greenwood, Stamford, CT), pp. 69–95.

Jusczyk, P., Luce, P. A., and Charles-Luce, J. (**1994**). "Infants' sensitvity to phonotactic patterns in the native language," J. Mem. Lang. **33**, 630–645.

Jusczyk, P. W. (**1985**). "On characterizing the development of speech perception," in Neonate Cognition: Beyond the Blooming, Buzzing Confusion, edited by J. Mehler and R. Fox (Erlbaum, Mahwah, NJ), pp. 199–229.

Jusczyk, P. W., Copan, H., and Thompson, E. (**1978**). "Perception by 2-month-old infants of glide contrasts in multisyllabic utterances," Percept. Psychophys. **24**, 515–520.

Kuhl, P. K. (**1979**). "Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories," J. Acoust. Soc. Am. **66**, 1668–1679.

Loebach, J. L., and Wickesburg, R. E. (**2006**). "The representation of noise vocoded speech in the auditory nerve of the chinchilla: Physiological

correlates of the perception of spectrally reduced speech," Hear. Res. **213**, 130–144.

Loizou, P. C. (**1998**). "Mimicking the human ear: An overview of signal processing techniques for converting sound to electrical signals in cochlear implants," IEEE Signal Process. Mag. **15**, 101–130.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., and Morgan, J. L. (**1999**). "Phonotactic and prosodic effects on word segmentation in infants," Cognit. Psychol. **38**, 465–494.

Munson, B., Donaldson, G. S., Allen, S. L., Collison, E. A., and Nelson, D. A. (**2003**). "Patterns of phoneme perception errors by listeners with cochlear implants as a function of overall speech perception ability," J. Acoust. Soc. Am. **113**, 925–935.

Niparko, J. K., Tobey, E. A., Thal, D. J., Eisenberg, L. S., Wang, N. Y., Quittner, A. L., Fink, N. E., and the CDaCI Investigative Team (**2010**). "Spoken language development in children following cochlear implantation," J. Am. Med. Assoc. **303**, 1498–1506.

Nittrouer, S., and Lowenstein, J. H. (**2010**). "Learning to perceptually organize speech signals in native fashion," J. Acoust. Soc. Am. **127**, 1624–1635.

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

Polka, L., and Werker, J. F. (**1994**). "Developmental changes in perception of nonnative vowel contrasts," J. Exp. Psychol. Hum. Percept. Perform. **20**, 421–435.

Roder, B. J., Bushnell, E. W., and Sasseville, A. M. (**2000**). "Infants' preferences for familiarity and novelty during the course of visual processing," Infancy **1**, 491–507.

Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. Roy. Soc. B: Biol. Sci. **336**, 367–373.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (**1996**). "Statistical learning by 8-month-old infants," Science **274**, 1926–1928.

Schvartz, K. C., Chatterjee, M., and Gordon-Salant, S. (**2008**). "Recognition of spectrally degraded phonemes by younger, middle-aged, and older normal-hearing listeners," J. Acoust. Soc. Am. **124**, 3972–3988.

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

Souza, P. E., and Rosen, S. (**2009**). "Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech," J. Acoust. Soc. Am. **126**, 792–805.

Stevens, K. N. (**1980**). "Acoustic correlates of some phonetic categories," J. Acoust. Soc. Am. **68**, 836–842.

Stevens, K. N. (**1983**). "Acoustic properties used for the identification of speech sounds," Ann. N.Y. Acad. Sci. **405**, 2–17.

Svirsky, M. A., Stallings, L. M., Lento, C. L., Ying, E., and Leonard, L. B. (**2002**). "Grammatical morphologic development in pediatric cochlear implant users may be affected by the perceptual prominence of the relevant markers," Ann. Otol. Rhinol. Laryngol. Suppl. **189**, 109–112.

Uhler, K., Yoshinaga-Itano, C., Gabbard, S. A., Rothpletz, A. M., and Jenkins, H. (**2011**). "Longitudinal infant speech perception in young cochlear implant users," J. Am. Acad. Audiol. **22**, 129–142.

Werker, J. F., Gilbert, J. H., Humphrey, K., and Tees, R. C. (**1981**). "Developmental aspects of cross-language speech perception," Child Dev. **52**, 349–355.

Werker, J. F., Shi, R., Desjardins, R., Pegg, J. E., Polka, L., and Patterson, M. (**1998**). "Three methods for testing infant speech perception," in *Perceptual Development: Visual, Auditory, and Speech Perception in Infancy*, edited by A. Slater (Psychology Press, New York), pp. 389–420.

Werker, J. F., and Tees, R. C. (**1984**). "Cross language speech perception: Evidence for perceptual reorganization during the first year of life," Infant Behav. Dev. **7**, 49–63.

Wetherford, M. J., and Cohen, L. B. (**1973**). "Developmental changes in infant visual preferences for novelty and familiarity," Child Dev. **44**, 416–424.

Xu, L., Thompson, C. S., and Pfingst, B. E. (**2005**). "Relative contributions of spectral and temporal cues for phoneme recognition," J. Acoust. Soc. Am. **117**, 3255–3267.

Xu, L., and Zheng, Y. (**2007**). "Spectral and temporal cues for phoneme recognition in noise," J. Acoust. Soc. Am. **122**, 1758–1764.