Using speech sounds to test functional spectral resolution in listeners with cochlear implants

Matthew B. Winn[a)]
Waisman Center and Department of Surgery, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, Wisconsin 53705

Ruth Y. Litovsky
Waisman Center, Department of Communication Sciences and Disorders and Department of Surgery, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, Wisconsin 53705

(Received 7 August 2014; revised 26 November 2014; accepted 20 January 2015)

In this study, spectral properties of speech sounds were used to test functional spectral resolution in people who use cochlear implants (CIs). Specifically, perception of the /ba/-/da/ contrast was tested using two spectral cues: Formant transitions (a fine-resolution cue) and spectral tilt (a coarse-resolution cue). Higher weighting of the formant cues was used as an index of better spectral cue perception. Participants included 19 CI listeners and 10 listeners with normal hearing (NH), for whom spectral resolution was explicitly controlled using a noise vocoder with variable carrier filter widths to simulate electrical current spread. Perceptual weighting of the two cues was modeled with mixed-effects logistic regression, and was found to systematically vary with spectral resolution. The use of formant cues was greatest for NH listeners for unprocessed speech, and declined in the two vocoded conditions. Compared to NH listeners, CI listeners relied less on formant transitions, and more on spectral tilt. Cue-weighting results showed moderately good correspondence with word recognition scores. The current approach to testing functional spectral resolution uses auditory cues that are known to be important for speech categorization, and can thus potentially serve as the basis upon which CI processing strategies and innovations are tested.
 [http://dx.doi.org/10.1121/1.4908308]

[MSS] Pages: 1430–1442

I. INTRODUCTION

Spectral resolution in hearing is the perceptual ability of a listener to distinguish between sounds that differ in pitch or other qualities in the spectral (i.e., frequency) domain. This aspect of hearing has clear implications for speech perception, as many speech sounds (e.g., /b/-/d/-/g/, /p/-/t/-/k/, most vowels) vary by spectral cues such as the frequency of formant peaks and other global spectral properties.

For people who use cochlear implants (CIs), spectral resolution is known to be degraded for multiple reasons. For example, there are a limited number of place-specific electrodes in the cochlea, resulting in poor specification of the spectral place code along the basilar membrane. Furthermore, electrical current fields generated by each electrode are known to overlap when using monopolar stimulation mode (which is the most common mode of stimulation). This results in electrical channel interaction, observable in various psychophysical tests (Boëx *et al.*, 2003; Abbas *et al.*, 2004). The interaction of current between electrodes, along with a host of other factors (e.g., history of deafness and variables relating to surgery and medical history) contributes to the common observation that listeners with CIs generally perform worse than listeners with normal hearing (NH) on tests of speech recognition and various psychoacoustic measures relating to spectral resolution. However, many of those psychophysical methods involve stimuli that are not speech, and that are not categorized in a speech-like manner. The objective of this study was to capitalize on the spectral properties of speech sounds and use them to measure the functional use of spectral cues by listeners with CIs. It was hypothesized that the use of natural speech cues in a categorization task would provide a "functional" test that has more direct correspondence to speech perception.

Measuring spectral resolution is important because one of the main goals in improving CI technology is the improvement of spectral resolution (cf. Bonham and Litvak, 2008), and several technological advances have emerged toward this goal. For example, speech processors can be made to steer current between neighboring electrodes, yielding a number of pitch percepts that far exceeds the number of physical electrodes (Firszt *et al.*, 2007). Electrical current from a single electrode can also be shaped by counter-weighted currents in adjacent electrodes in order to increase the specificity of neural stimulation (Bierer, 2007; Srinivasan *et al.*, 2010) Benefit from these techniques is sometimes difficult to ascertain; some listeners have been reported to prefer current focusing despite showing no demonstrable increase in word recognition performance (Mens and Berenstein, 2005). There is a need for sensitive tests to quantify the benefit that these listeners are reporting. It is likely the case that conventional speech recognition testing is not a sensitive-enough tool to capture such benefits

a)Author to whom correspondence should be addressed. Electronic mail: mwinn83@gmail.com

because it is affected by a number of factors other than spectral resolution.

A. Previous measures of spectral resolution in CI listeners

The most popular methods of testing spectral resolution for CI listeners are non-linguistic tasks such as electrode discrimination (Nelson *et al.*, 1995; Zwolan *et al.*, 1997), electrode pitch ranking (Donaldson and Nelson, 2000), and spectral ripple discrimination (Henry *et al.*, 2005; Won *et al.*, 2007; Jones *et al.*, 2013). Many of these tests have shown a relationship between spectral resolution and speech perception, but each test is limited in important ways that are addressed in the current study. First, spectral modulations in non-linguistic stimuli are not necessarily representative of the most important spectral modulations in speech (Singh and Theunissen, 2003). For non-tonal languages, high-density spectral modulations such as harmonic partials are likely to be less important for intelligibility than low-density spectral modulations that comprise formant structure. A study by Saoji *et al.* (2009) confirmed the importance of broadly spaced spectral modulations for CI listeners. Spectral ripple stimuli at best do not offer a frequency-specific profile of a listener's abilities, and at worst test spectral "edges" that are thought to be only negligibly important in the transmission of speech information (Azadpour and McKay, 2012). Finally, both spectral ripple tests and psychophysical tests of electrode discrimination fail to capture the essential nature of speech perception as a process of *categorization* rather than discrimination (Holt and Lotto, 2010). Some criticism of spectral ripple stimuli have been addressed by rigorous parallel examination of electrical current spread measures (Jones *et al.*, 2013) and the use of drifting phase relationships between ripple components to avoid spectral edge-detection cues (Aronoff and Landsberger, 2013). Despite the promise of such efforts, the current study takes a different approach using modified natural speech sounds in hopes of establishing a firmer connection to speech identification abilities more generally.

B. Testing for spectral resolution using speech sounds

Speech in its natural form is usually not a good stimulus for testing spectral resolution because they contain a large number of co-varying cues, many of which are not spectral in nature. For example, there are 24 different acoustic cues that distinguish the English fricative sounds (e.g., /s/, /f/, /z/, etc.) (McMurray and Jongman, 2011), and 16 different cues that distinguish the /b/ and /p/ sounds in word-medial position (Lisker, 1986). Some of these cues involve durational differences, or differences in the amplitude envelope of the sounds, or of adjacent sounds. Thus, even when perception of a target sound is "correct," it does not necessarily imply that a spectral cue was recovered; compensatory behavior in the temporal domain can allow listeners to appear to have good speech identification performance despite atypical perception of acoustic cues (Winn *et al.*, 2012). In spite of this confound, there is good correspondence between spectral resolution and errors on speech contrasts that are *primarily* spectral in nature (Xu *et al.*, 2005). In this study, the consonant place contrast was exploited toward the goal of developing a test that was both focused on spectral resolution and was also challenging enough to capture differences among listeners of wide-ranging abilities.

Despite the multidimensional nature of speech acoustics, there are some patterns that clearly emerge when exploring the impact of spectral resolution. In general, consonant *place of articulation* (i.e., the distinction between word pairs like "big" and "dig" or "take" and "cake") is the consonant feature most heavily driven by spectral cues, such as consonant burst spectrum and formant transitions. Understandably, place of articulation perception is particularly difficult for people with hearing impairment (Bilger and Wang, 1976), including people with CIs (Munson *et al.*, 2003), and is a particularly difficult feature to recover when spectral quality is experimentally degraded for listeners with NH (Friesen *et al.*, 2001; Xu *et al.*, 2005). When conducting studies with NH individuals, resolution in the spectral domain can be explicitly controlled using noise vocoders (Shannon *et al.*, 1995; Friesen *et al.*, 2001; Xu *et al.*, 2005), sine vocoders (Dorman and Loizou, 1997), Gaussian envelope tone vocoders (Goupell *et al.*, 2013), and also with other methods used to achieve spectral smearing (ter Keurs *et al.*, 1993). In these studies, place of articulation consistently emerges as a difficult feature to recover.

The impact of particular cues on speech sound categorization can be explored in studies of perceptual cue weighting. Typically in such studies, multiple cues to a speech sound contrast are orthogonally manipulated, and listeners' categorization responses can demonstrate that they place more perceptual weight on some cues over others, or that change in one dimension can counteract change in another dimension (Repp, 1982). For the purpose of this study, the relevant phonetic cues are spectral in nature; the philosophy behind this approach is that better functional spectral resolution should permit more efficient use of spectral cues for speech contrasts. Presumably, if listeners with NH demonstrate heavy reliance on specific spectral cues in their perception of speech contrasts, their cue-weighting patterns can serve as the basis for evaluating the performance of listeners with CIs.

Isolation of specific cues in speech can be challenging. Synthetic speech permits rigorous control, but there are some issues that limit the utility of this approach. For instance, synthetic speech is consistently less intelligible than natural speech (Greene *et al.*, 1986), and there are several published studies suggesting that listeners perceive spectral and temporal cues in synthetic speech in a qualitatively different way than in natural speech (Walsh and Parker, 1984; Hillenbrand *et al.*, 2000; Nittrouer, 2004). Across these studies, spectral cues like formant structure and formant transitions were weighted less heavily in synthetic speech compared to natural (or modified natural) speech, presumably because of limitations in the ability to synthesize natural-sounding voices, or limitations in assessing all of the relevant details of the speech spectrum. As formant transitions and formant structure are precisely the cues that are

relevant for the current discussion, traditional synthetic speech is not an ideal probe for spectral resolution.

C. Cue weighting as a metric of spectral resolution

Acoustic phonetic cue weighting is an ideal approach to probe functional spectral resolution for speech because it is driven by the functional acoustic units thought to underlie speech perception, and it can potentially resolve subtle differences between listeners of different hearing abilities who may show equivalent performance on word recognition. In short, those listeners who are best able to resolve spectral cues should demonstrate higher perceptual weighting for a spectral speech cue, given the dominance of that cue for the tested contrast. It should be noted that this measurement reflects the functional ability of listeners to perceive and *use* a cue, rather than an absolute psychophysical ability to resolve the cue. This concept of functional spectral resolution is qualitatively different than the traditional idea of absolute frequency discrimination, but is arguably the basis of auditory linguistic category formation.

Phonetic cue weighting is thus far an underutilized window into the auditory processing abilities of listeners with hearing impairment. Previous literature suggests that phonetic cue weighting is affected by hearing loss (Alexander and Kluender, 2009; Revoile *et al.*, 1985), masking noise (Wardrip-Fruin, 1985; Winn *et al.*, 2013a), spectral bandwidth (Winn *et al.*, 2013a), and cochlear implantation (Winn *et al.*, 2012, 2013b). Using NH listener cue-weighting patterns as a model for optimal performance, cue-weighting analysis can be used to evaluate how well CI listeners recover and use specific acoustic components of speech and, in the future, whether some intervention (e.g., current focusing, current steering, altered frequency-electrode allocation, etc.) improves one's ability to recover those spectral cues.

Place of articulation for stop consonants has been identified as a particularly important and difficult contrast for listeners with CIs. This contrast is cued in large part by formant transitions and spectral tilt (i.e., relative balance of high- and low-frequency information in the spectrum) of the burst and vowel onset relative to the central part of the vowel (Alexander and Kluender, 2008). Blumstein *et al.* (1982) described "gross spectral shape" as an amalgamation of both of these cues. Formant transitions and spectral tilt are ideal cues to test spectral resolution because virtually all listeners, whether they have good or poor spectral resolution, retain mechanisms to recover the place of articulation "information"; those with excellent spectral resolution should conform to the "typical" NH pattern of relying upon formant transitions, while those with poorer spectral resolution are likely to fall back on the secondary cue of spectral tilt, which is accessible even with poor resolution, and is used more heavily by listeners with hearing impairment (Alexander and Kluender, 2009). The relative reliance on the formant transition cue can thus be used as a proxy index of spectral resolution.

II. METHOD

A. Participants

Participants included 19 listeners with bilateral cochlear implants (BiCIs), whose demographics are listed in Table I. All BiCI users were native speakers of American English, and all but three were post-lingually deafened. There were also 10 listeners with NH (ANSI, 2010), between the ages of 18 and 31, who were all native speakers of American English. There was a substantial age difference between the CI and NH listener groups, mainly due to the availability of the test populations.

B. Stimuli

Speech stimuli consisted of modified natural speech tokens that were spoken by a native speaker of American English. There were three classes of sounds, described below.

1. /ba/-/da/ continuum

The /ba/-/da/ continuum featured orthogonal manipulation of formant transitions and spectral tilt at the onset of the syllable. These two parameters were controlled so that there was a cue that varied within a narrow frequency range (i.e., the second formant varied between roughly 1000 and 1800 Hz) and a cue that varied across a wide frequency range (i.e., the spectral tilt cue varied over the range between roughly 800 and 6000 Hz). The creation of these stimuli was a multi-stage process that is illustrated in a detailed step-by-step diagram in Fig. 1, and is described in the following paragraphs.

First, a continuum varying by formant transitions was created using a modification of the basic linear predictive coding (LPC) decomposition and re-synthesis procedure in Praat (Boersma and Weenink, 2013). A naturally produced /ba/ token spoken by a male talker was first downsampled to 10 000 Hz to facilitate accurate estimation of 12 LPC coefficients below 5000 Hz. The downsampled sound was inverse filtered by the LPC, which is a common way to eliminate formant peaks in the sound, in order to yield a residual sound (i.e., "source") with a speech-like sloping spectrum that can be filtered with a different formant structure. Formant contours were extracted from the original /ba/ and /da/ utterances, and six intermediate contours were made via linearly interpolation. Figure 2 shows a schematic illustration of formant contour parameters across the 8-step continuum. The center frequency of $F2$ at onset ranged from 1000 to 1800 Hz, which is consistent with published reports of synthesized /ba/ and /da/ sounds (Francis *et al.*, 2008). The consonant release burst was also filtered by the onsets of the formant contours and thus complemented the formant transitions.

An obligatory limitation of the aforementioned LPC source-filter procedure is the loss of high-frequency energy above 5000 Hz (because of the 10 000 Hz downsampling). To restore this high-frequency energy, each step of the continuum of speech sounds was further low-pass filtered at 3500 Hz and added with the original /ba/ signal that was

TABLE I. Demographic information for CI users in this study.

Number	Code	Sex	Age	Years CI exp.	Years BiCI exp.	Implant type (Left)	Implant type (Right)	External processor (Left)	External processor (Right)
1	IAJ*	F	67	16	9	CI24M	CI24R (CS)	N5	N5
2	IBF	F	62	7	5	Freedom Contour Adv.	Freedom Contour Adv.	Freedom	Freedom
3	IBK	M	72	9	3	Nucleus 24 Contour	Freedom Contour Adv.	Freedom	N5
4	IBM	F	59	7	3	CI512 (N5)	Freedom Contour Adv.	N5	N5
5	IBN*	M	66	13	3	Freedom Contour Adv.	Nucleus 24 Contour	Freedom	Freedom
6	IBO	F	46	6	3	CI512 (N5)	Freedom Contour Adv.	N5	Freedom
7	IBR	F	57	9	5	CI512 (N5)	Freedom Contour Adv.	N5	Freedom
8	ICA	F	52	10	3	Freedom Contour Adv.	Freedom Contour Adv.	N5	N5
9	ICB	F	61	8	6	Freedom Contour Adv.	Nucleus 24 Contour	Freedom	N5
10	ICD*	F	54	9	3	Freedom Contour Adv.	Nucleus 24 Contour	Freedom	Freedom
11	ICF	F	70	1	1	CI512 (N5)	CI512 (N5)	N5	N5
12	ICG*	F	50	9	9	Freedom Contour Adv.	Freedom Contour Adv.	N5	N5
13	ICJ	F	63	3	1	CI512 (N5)	CI512 (N5)	N5	N5
14	ICM	F	59	1	1	CI512 (N5)	CI512 (N5)	N5	N5
15	ICO	F	32	1	1	CI512 (N5)	CI512 (N5)	N5	N5
16	ICP*	M	50	6	3	Freedom Contour Adv.	Freedom Contour Adv.	N5	N5
17	ICQ*	M	19	15	1	Freedom Contour Adv.	Nucleus 24 straight	N5	N5
18	ICR	M	60	3	2	CI512 (N5)	CI512 (N5)	N5	N5
19	ICV	M	58	6	5	Freedom Contour Adv.	Freedom Contour Adv.	N5	N5

Note: Asterisks indicate listeners who were deafened before age 7.

high-pass filtered above 3500 Hz; both the low- and high-pass filters were specified with a symmetrical 500 Hz roll-off using the Hann filter in Praat. Thus, the full spectrum of frequency energy was preserved in the final stimuli. Both the low- and high-frequency components had perfect phase synchrony because they operated on sounds with the same original time domain. The high-frequency energy from the original /ba/ token was constant across the entire formant continuum, but added considerable naturalness of the sound. All steps in the formant continuum were equalized for spectral tilt measured from the first to the fourth formant at syllable onset, defined as the burst and the first 80 ms of the vowel.

Spectral tilt was systematically modified in five steps within each step in the formant continuum, using a filter that amplified or attenuated frequency energy above 800 Hz via logarithmic multiplication of the amplitude spectrum with varying slope across the frequency range. Across the spectral tilt continuum, spectra were maximally distinct at *F*4 (3300 Hz) and tapered to equivalent levels at 6000 Hz as follows: The output of the multiplication filter was low-passed at 8000 Hz with a symmetrical 2000 Hz roll-off filter and combined with complementary high-pass filtered energy from the third (neutral) step in the continuum. The result of these complementary filters was that the filtered sound and original sounds tapered together at 6000 Hz; this frequency was an anchor above which all spectral envelopes were congruent across the spectral tilt continuum. This complementary high/low-pass filtering procedure was similar to that done for the formant continuum, and was necessary to restrict the impact of spectral tilt to the frequency range generally observed in the analysis of this acoustic cue.

Following the spectral tilt filtering, the sounds were blended into the original vowel from /ba/ using an 80 ms linear cross-fading transition window beginning at the consonant release burst; at 80 ms relative to the release, all stimuli were congruent, and were composed simply of the vowel nucleus from the original unedited "ba" utterance. Thus, the tilt modification affected only the consonant release/vowel onset, resulting in a *dynamic* cue of *relative* spectral tilt, in accordance with measurements of production and effects on perception (Alexander and Kluender, 2008, 2009). The uniform vowel offset also neutralized any late-occurring cues to consonant identity. Refer to Fig. 3 for a detailed step-by-step diagram of the spectral tilt modification procedure for this stimulus set.

By combining the cross-fading procedure with the aforementioned inclusion of high-frequency energy from the original /ba/ recording, the manipulated portions of the stimuli were essentially "boxed in" by the natural signal in both time and frequency. Each stimulus began with a uniform 100 ms segment of prevoicing from /ba/, which was deemed to be perceptually indistinct from that for /da/. Figure 4 illustrates the spectral tilt parameters described here, as well as the spectrum of the vowel nucleus, which was constant across all stimuli in this set.

2. /ʃ/-/s/ ("sha-sa") continuum

There were seven sounds in the /s/ and /ʃ/ category, comprised of a 7-step continuum whose endpoints were natural /s/ and /ʃ/ sounds. Intermediate steps were created by graduated blending of these signals. Consistent with the naturally-produced signals, the fricative spectrum peaks varied in terms of relative amplitude of spectrum peaks within the fricative noise. CI listeners generally do not show difficulties categorizing /s/ and /ʃ/ sounds (Winn *et al.*, 2013b),

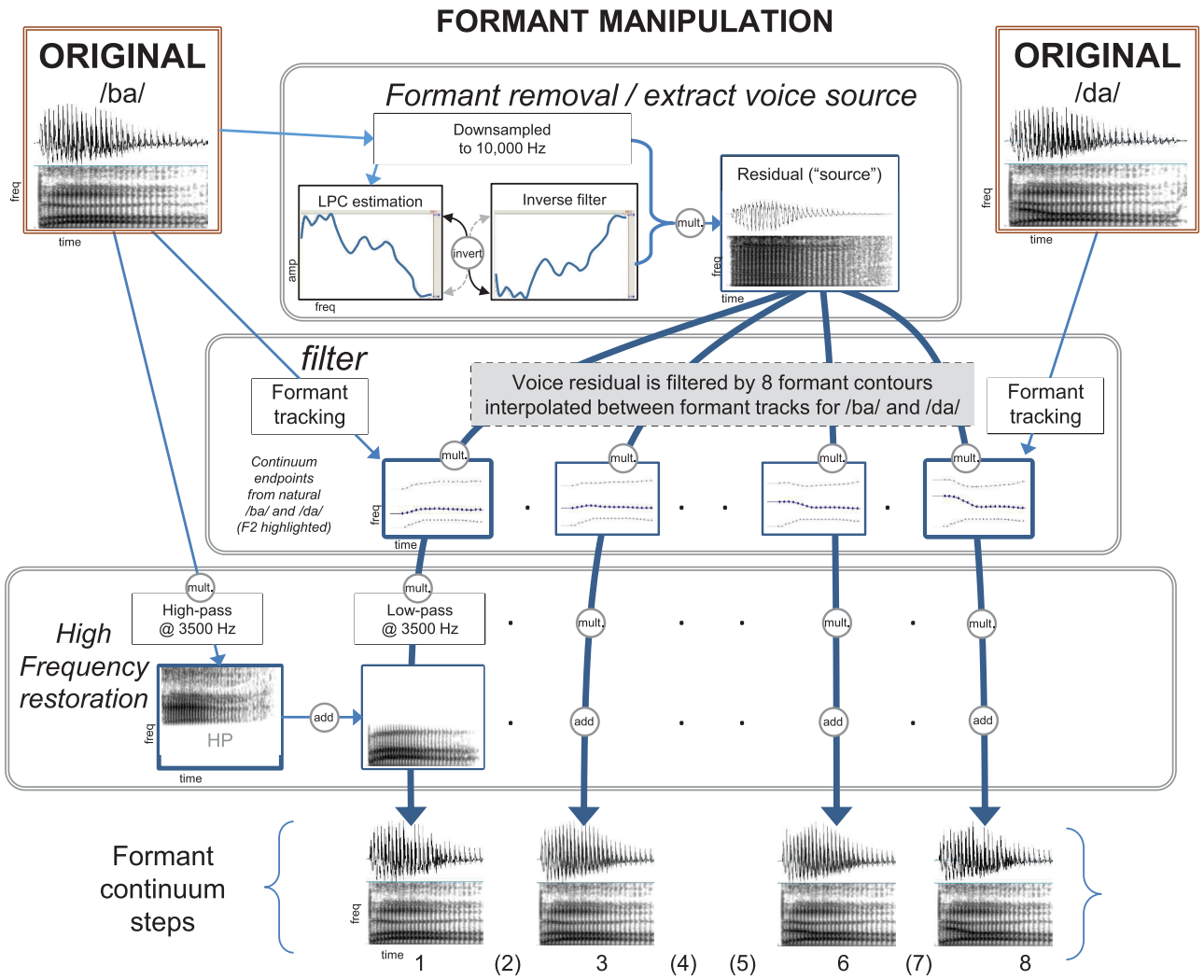

FIG. 1. (Color online) Flowchart of signal processing procedures implemented in the Praat software to create a continuum of speech sounds varying in formant structure. Formant contours from /ba/ and /da/ were used as endpoints between which 6 other contours were interpolated to create *FormantGrid*s. Each of those FormantGrids was used to filter a residual sound that was created by removing the formant peaks from the original /ba/ sound using LPC inverse filtering. After the filtering, each resulting sound was low-pass filtered and added with high-pass frequency energy from the original /ba/ sound.

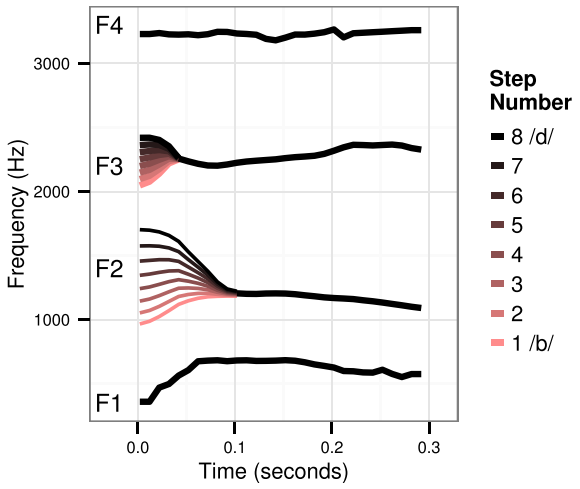

FIG. 2. (Color online) Schematic of formant contours for /ba/-/da/ stimuli. Lowest *F*2 onset (lighter-colored) lines are most /ba/-like, and highest *F*2 onsets (black) are most /d/-like. *F*1 and *F*4 were congruent across all steps of the continuum. *F*2 was congruent for all steps starting at 0.1 s post-onset, and *F*3 was congruent for all steps starting at 0.4 s post-onset.

so these stimuli were not used for assessment of any specific abilities.

3. /ra/ and /la/ sounds

The two remaining choices in the 6-alternative were /ra/ and /la/, which were unaltered natural recordings of these syllables.

4. Monosyllabic words

Open-set monosyllabic word recognition was tested using a set of 210 monosyllabic words of consonant-vowel-consonant structure sampled from the Maryland CNC clinical corpus. After each word was presented, the participant would type his/her response and click a button to proceed to the next word. Virtually all of the words used were listed in the Hoosier Mental Lexicon (Nusbaum *et al.*, 1984), which describes the familiarity and frequency of the words; of those words, 97% were designated as having excellent familiarity (at least 6.5 on a 7-point scale), including those that occur only rarely (e.g., "kite"). These words had an average

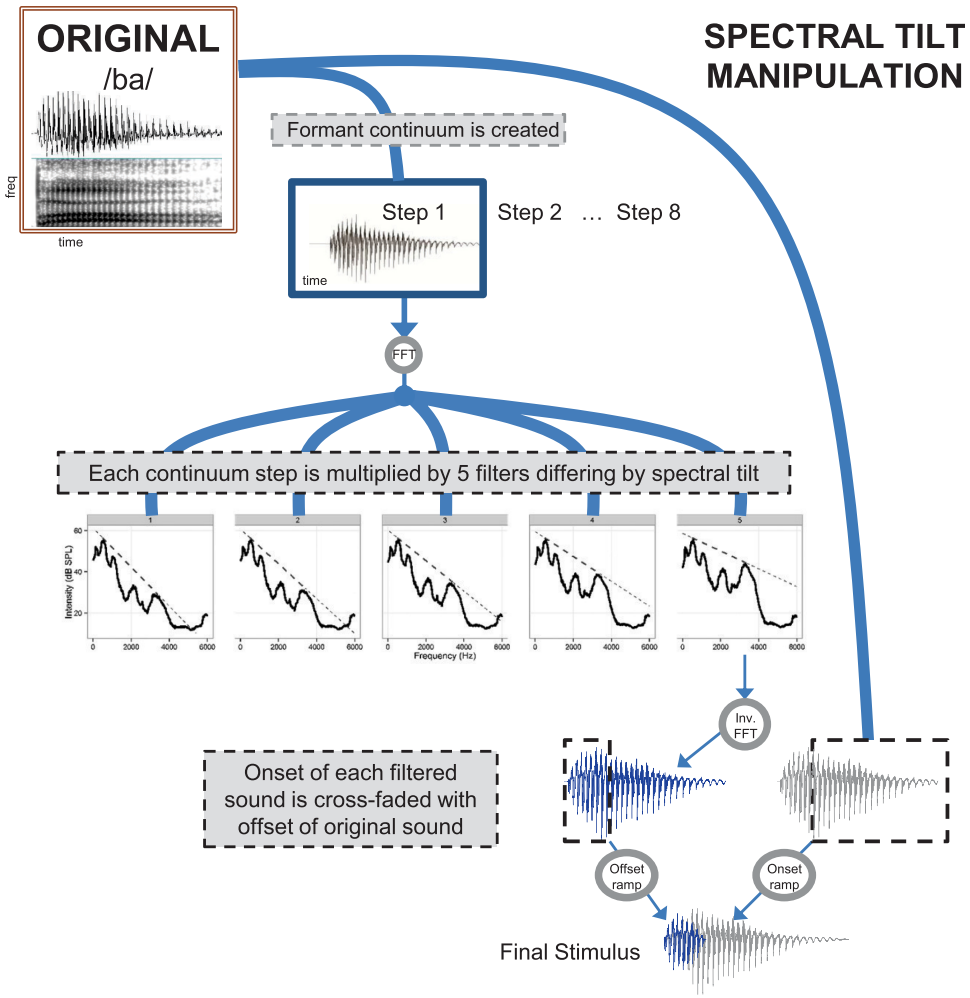

FIG. 3. (Color online) Flowchart of signal processing procedures implemented in the Praat software to create a continuum of speech sounds varying in spectral tilt at syllable onset. Each of five linearly sloped filters was applied to each step in the formant continuum. The result of that filtering was cross-faded into a uniform vowel nucleus from the original /ba/ sound that was not filtered. As a result, only the onset of the syllable was filtered, and it gradually morphed back into the unfiltered vowel.

of 20 lexical neighbors (ranging from 3 to 31), defined by potential real-word confusions resulting from a phoneme substitution, addition, or deletion. For NH listeners, 50 words were tested in each listening condition; CI listeners heard only 50 words. Ten words were used as a practice set to familiarize each participant with the procedure.

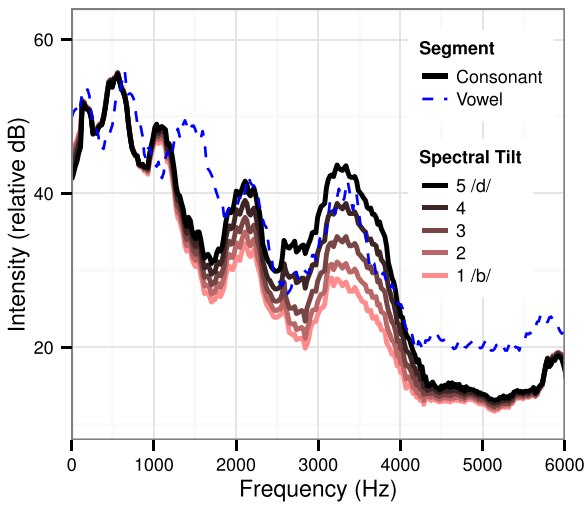

FIG. 4. (Color online) Continuum of spectral tilt, with formant structure held constant. Steepest tilt (lighter-colored) is most /ba/-like, and shallowest tilt (black) is most /da/-like. All continuum steps has congruent spectra for frequencies below 800 Hz and above 6000 Hz.

C. Noise vocoding

It is essential that a test of spectral resolution be sensitive to known differences in spectral resolution. To create such differences, spectral resolution was explicitly controlled for listeners with NH using two vocoder conditions designed to simulate the processing strategy used by the CI listeners. In contrast to previous literature that controlled spectral resolution in vocoders by varying the number of discrete spectral channels (cf. Shannon *et al.*, 2004), we modified the approach used by Litvak *et al.* (2007); the number of analysis and carrier channels was kept constant, while spectral resolution was controlled by applying varying filter slopes on the carrier filters so as to simulate varying amounts of "current spread." Using the AngelSim software (Fu *et al.*, 2013), speech signals were divided into 22 spectral analysis filters whose corner frequencies were linearly interpolated in cochlear space using the Greenwood function (Greenwood, 1990). Absolute lower- and upper-frequency boundaries were set to equal those used by the CI listeners in this study (188 and 7938 Hz, respectively; these are the corner frequencies of the Cochlear Nucleus speech processors).

In each analysis window, the 8 frequency channels with greatest energy (following pre-emphasis) were selected for synthesis. Figure 5 illustrates vocoder channel peak-picking for a segment of a vowel sound. Local peaks in the original sound spectrum excite the 22 vocoder filters, out of which 8

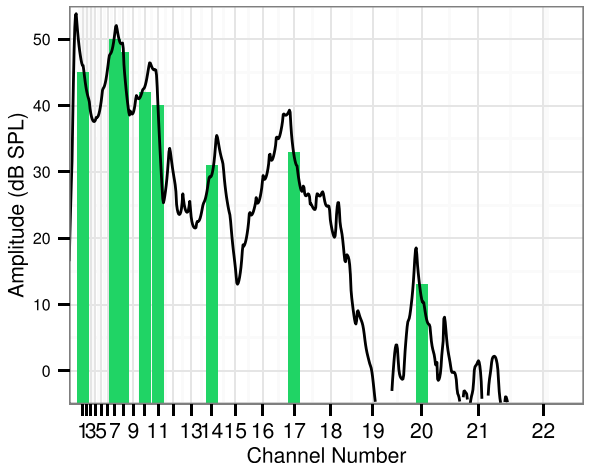

FIG. 5. (Color online) Illustration of a smoothed spectrum (black line) divided into 22 cochlea-spaced channels, out of which 8 peaks are picked for synthesis. Such peak-picking was performed on rolling time windows and represents the style of spectrum processing commonly found in CI speech processors.

synthesis channels represent the input sound. Selection of those 8 peaks is motivated by the most common setting for the Advanced Combination Encoder processing strategy, which was used all 19 of our CI listeners. This strategy essentially reduces the number of electrodes in order to alleviate some of the interactions between channels; wider channel carriers increase channel interaction and thus cause more spectral degradation. Settings in the AngelSim software were adjusted to produce carrier band filter slopes of 50 dB/octave (easy) and 9 dB/octave (difficult), representing low and high degrees of channel interaction, respectively.[1]

D. Procedure

Listeners completed a one-interval 6-alternative forced choice syllable identification task. On each trial, one stimulus was presented, and listeners responded using one of the following six choices: /ba/, /da/, /sa/, /ʃa/, /la/, /ra/. The crucial contrast was between /ba/ and /da/, and the other four sounds were included merely to give variety to the stimuli, in order to avoid artificially heightened sensitivity to manipulated stimulus dimensions. Listeners were not made aware that any syllables were manipulated. Because of the salient envelope differences between the three different stimulus sets, the number of mis-matched sound category responses was negligible.

All listeners began with a short practice session to familiarize them with the experiment interface, the speech sounds, and the procedure. Three blocks of randomized stimuli were played in each condition; CI listeners only heard the unprocessed speech condition, while NH listeners also heard the two extra vocoder conditions. For NH listeners, the first block was always an unprocessed speech (i.e., non-vocoded) set, and there was a short practice set of vocoded stimuli presented before the first vocoded test block. Stimuli within each block were randomized. The crucial /ba/-/da/ contrast was slightly over-represented in the stimulus list (40% of trials, instead of 33%), so that more data were collected on the specific measure of interest. Listeners also participated in a test of open-set monosyllabic word recognition in the same testing conditions. Fifty words were randomly selected from a corpus of 415 total words for each listener in each condition. Upon hearing each word, participants typed their responses on a computer. No feedback was offered during any component of testing. If a participant asked for feedback, the standard response was: "There are no correct or incorrect answers to this test; we are looking to see whether these pronunciations are judged consistently, no matter what you think they are."

E. Analysis

The main group analysis was restricted to responses to syllables in the /ba/-/da/ continuum. Participants' identification responses were modeled using generalized linear (logistic) mixed-effects regression (GLMM) using the lme4 package (Bates *et al.*, 2014) in the R software environment (R Core Development Team, 2014). The binomial outcome variable was perceived place of articulation (ba or da). The fixed-effect predictors included continuous dimensions of formant transitions and spectral tilt (coded as centered continuum step), and, for NH listeners, the vocoder condition. The random-effects structure included random intercepts and random slopes for both cues for each participant in each condition.

In addition, individual analysis was performed for each listener in each condition, without any random effects structure. Consistent with Agresti (2002) and Barr (2008), individual psychometric functions were modeled using the *Empirical Logit* transformation, which uses a logit linking function but more gracefully accommodates situations in which odds ratios approach negative or positive infinity. This method was chosen because individual datasets are much sparser than group data sets, and are more likely to yield such undefined odds ratios at continuum endpoints, or in situations of perfect separation (i.e., when there is a perfect step function ranging from 0 to 1 with no continuum steps yielding intermediate proportions). The empirical logit transformation adds a 0.5 adjustment factor, which corrects for undefined ratios and scales to the granularity of individual datasets, including those with unequal numbers of observations. Whereas the classical logit transformation is defined as $logit(\text{hits, trials}) = \log(\text{hits/misses})$, the empirical logit is defined as $elogit(\text{hits, trials}) = \log([\text{hits} + 0.5]/[\text{misses} + 0.5])$. Because the empirical logit is a linear approximation of data whose variance structure is non-linear, weights were applied to observations based on their reliability, following the technique described by Mirman (2014).

III. RESULTS

A. Group results

The top panels of Fig. 6 show the difference in the psychometric function morphology across listener groups in response to the formant cue; a well-defined steeply sloping function was observed for NH listeners in the unprocessed condition, and functions grew shallower as spectral resolution was degraded and is shallowest for listeners with CIs.

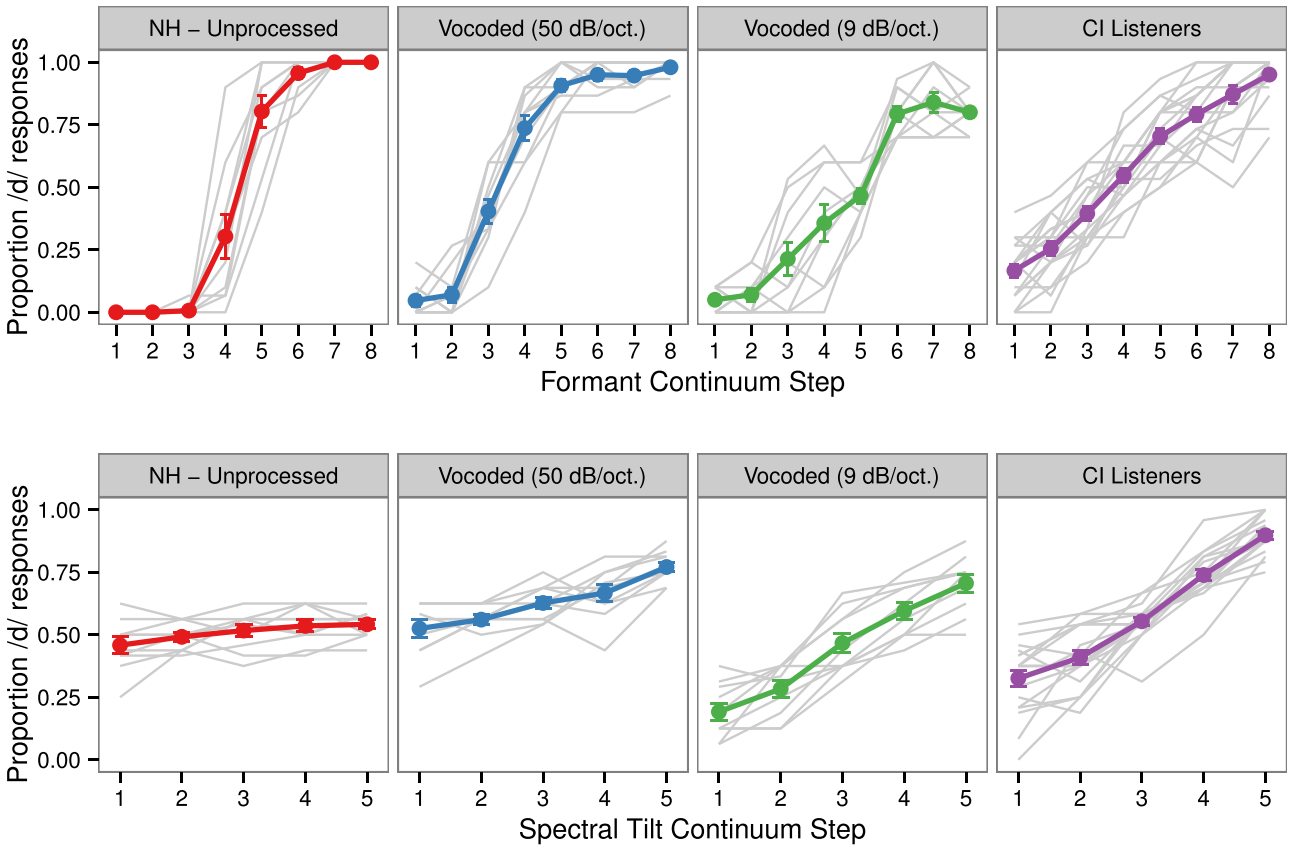

FIG. 6. (Color online) Psychometric functions for each listener group reflecting the proportion of /ba/ perceptions along the continuum of formant cues (upper panel) and spectral tilt cues (lower panel) as they ranged from /ba/ to /da/.

The opposite pattern emerged in response to the spectral tilt cue, where NH listeners had the shallowest functions and the CI listeners had steeply sloping functions.

Group results were modeled using the aforementioned GLMM; higher model coefficients correspond to steeper slopes for the psychometric functions. Higher coefficients for the formant cue were used as indices of good spectral resolution, and higher coefficients for spectral tilt were taken as indices of poor spectral resolution.

The effect of the formant cue was strongest for NH listeners in the unprocessed condition. Compared to that condition, the effect was significantly smaller for the vocoded signals with 50 dB/octave filter slopes ($z = -5.93$; $p < 0.001$) and significantly weaker for the vocoded signals with 9 dB/octave filter slopes ($z = -6.7$; $p < 0.001$). Formant cue effects in the vocoded conditions approached, but did not reach, a significant difference from each other ($z = -1.774$; $p = 0.08$). In general, the effect of the formant cue decreased as spectral resolution grew poorer. The effect of the formant cue for CI listeners was significantly smaller than that for NH listeners in the unprocessed condition ($z = -6.26$; $p < 0.001$). CI listeners' use of the formant cue was not significantly different from that observed in NH listeners in the 50 dB/octave condition ($z = -1.45$; $p = 0.15$) nor from the 9 dB/octave filter condition ($z = 0.31$; $p = 0.76$).

The effect of spectral tilt across conditions generally showed a pattern opposite to that of the formant cue; as spectral resolution was degraded, reliance on the spectral tilt cue increased. For NH listeners, the effect of spectral tilt for the unprocessed condition was not significantly different from that for the vocoded signals with 50 dB/octave filter slopes ($z = -0.9$, $p = 0.32$). The effect of spectral tilt was significantly stronger for the vocoded signals with 9 dB/octave filter slopes ($z = 4.73$; $p < 0.001$), compared to the unprocessed condition. The effect was also significantly different across two vocoder conditions ($z = 4.55$; $p = 0.001$). The effect of the spectral tilt cue for CI listeners was significantly stronger than that for NH listeners in the unprocessed condition ($z = 5.46$; $p < 0.001$), and NH listeners in the 50 dB/octave filter condition ($z = 5.47$; $p = 0.001$). The effect of spectral tilt for CI listeners was not significantly different from that for NH listeners in the 9 dB/octave filter condition ($z = -0.58$; $p = 0.56$).

Perception of the "filler" syllables was generally excellent for all listeners across all conditions, with the exception of the /ra/ syllable heard by NH listeners in the 9 dB/octave vocoder condition. Table II shows summary statistics of each listener group for each of the four syllables. For the /s/ and /ʃ/ phonemes, the endpoints of the 7-step continuum were used as proxy "canonical" productions and were evaluated for accuracy. Psychometric functions along the /s/ - /ʃ/ continuum were extremely similar across listener groups and vocoder conditions, and are thus not displayed here.

B. Individual results

Individual response curves were modeled using the empirical logit estimation method, which produces model

TABLE II. Accuracy scores for four phonemes other than /b/-/d/.

Phoneme	NH Unprocessed	Vocoded (50 dB/oct.)	Vocoded (9 dB/oct.)	CI Listeners
/ʃ/	92.5	87.4	99.5	98.0
	(16.4)	*(29.8)*	*(1.7)*	*(4.2)*
/s/	100.0	99.7	94.7	85.5
	(0.0)	*(1.1)*	*(14.4)*	*(28.1)*
/r/	99.7	99.6	77.9	95.7
	(0.6)	*(1.3)*	*(30.2)*	*(7.9)*
/l/	99.5	97.2	93.8	93.2
	(1.3)	*(6.8)*	*(9.9)*	*(9.2)*

Note: /ʃ/ and /s/ scores are based on identification of endpoints of the 7-step continuum. Numbers in parentheses represent 1 standard deviation.

coefficients that can be interpreted the same as those described for the group model above. Figure 7 illustrates the levels and degree of variability across listener groups for each of the acoustic cues in this experiment. The values in this figure faithfully reflect the morphological differences observed in the psychometric functions, and give quantification to those trends. Consistent with the group model, individuals with NH show highest logit coefficients for the formant cue when speech signals were not vocoded, and those coefficients became smaller with successive degradations in spectral resolution. CI listeners showed a larger amount of variability, but their collective responses were similar to those observed from NH listeners in the 9 dB/octave vocoder condition. The opposite trend emerged for the spectral tilt cue.

Three of the listeners had roughly 1 year of experience with electric hearing, while the others had substantially more. A *post hoc* analysis excluding these three recently implanted listeners yielded results that were virtually indistinguishable from the overall group results reported here; there were no changes in effect directions or ordering between groups, and the largest z-value change was less than 0.03. Analysis of only the single-year implant users yielded a formant cue coefficient that was slightly higher than that of the longer-implanted group, likely because of the influence of participant "ICO," who demonstrated excellent use of that phonetic cue.

1. Cue trading

There was a highly significant negative correlation between logistic coefficients for the formant and spectral tilt cues ($r^2 = 0.79$; $p < 0.001$), indicating a cue-trading relationship that is illustrated in Fig. 8. As the effect of the formant cue was diminished by poor spectral resolution, the effect of the spectral tilt cue increased, consistent with the notion that these are complementary cues that are utilized in different ways depending on spectral resolution. Data from the different listener groups are roughly separated into groups along the regression line, suggesting that both measurements provide complementary evidence that phonetic cue weighting can be used as an index of spectral resolution.

2. Relationship to word recognition

For CI listeners, there was a significant correlation between word recognition scores and formant cue coefficient ($r^2 = 0.27$; $p = 0.02$). The relationship between word recognition and spectral tilt coefficient failed to reach significance for CI listeners ($r^2 = 0.04$; $p = 0.41$), although it approached significance when two listeners (Participants "ICV" and "ICQ") were excluded from analysis. For NH listeners, neither correlation reached significance (formant: $p = 0.92$; spectral tilt $p = 0.75$), likely because of a ceiling effect for word recognition leading to negligible variability across conditions. Figure 9 illustrates the relationship between the two

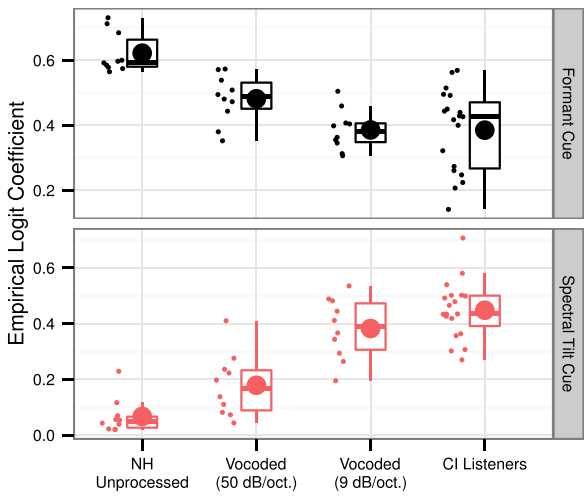

FIG. 7. (Color online) Boxplots showing empirical logit coefficients for individuals in each listening condition. The upper panel shows coefficients for the formant cue, and the lower panel shows coefficients for the spectral tilt cue. Lower and upper edges of boxes reflect the 25% and 75% interquartile range. Upper and lower whiskers reflect the median ±1.58 * the interquartile range/square root (n). Median levels are represented by the horizontal line crossing each box, and mean levels are represented by the large circles. Individual levels are represented by small circles.

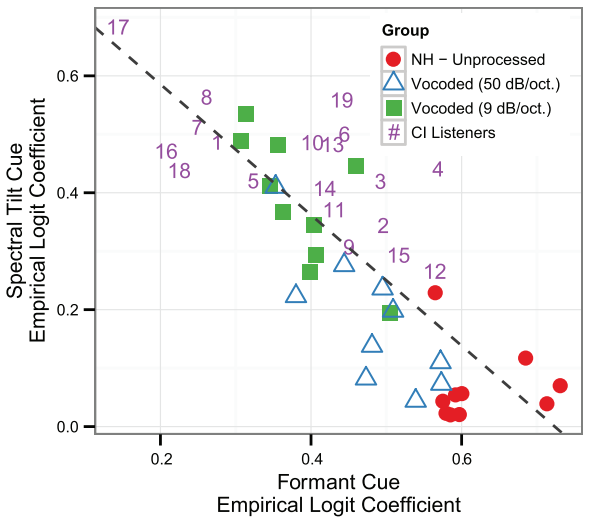

FIG. 8. (Color online) Scatterplot of empirical logit coefficients for formant cues (x axis) and spectral tilt cues (y axis) for each listener group. NH listeners in different conditions are represented by different shapes, while CI listeners are represented by unique subject numbers that correspond to the demographic table (Table I). Across all listener groups, higher formant cue coefficients are typically accompanied by lower spectral tilt coefficients, and vice versa.

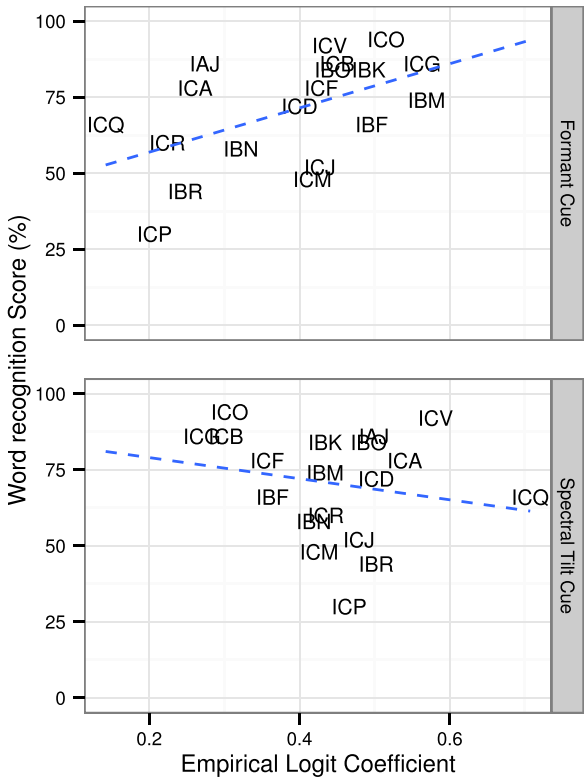

FIG. 9. (Color online) Correlations between individual participants' scores for word recognition and logistic coefficients for formant cues (left panel) and spectral tilt cues (right panel). Individual data points are jittered vertically within 3% to aid visibility.

cue coefficients and word recognition for CI listeners only. It can be seen that elevated formant cue coefficients are associated with higher word recognition scores. Higher spectral tilt coefficients are somewhat associated with lower word recognition scores, but this relationship is less clear. While word recognition is unsurprisingly affected by spectral resolution, it must be noted that this relationship was not expected to be especially strong, since words can be recognized using a variety of cues even when spectral cues are impoverished (Xu *et al.*, 2005; Winn *et al.*, 2012), implying that intelligibility performance on conventional word recognition tests are not adequate to measure spectral resolution. However, those listeners who are more adept at utilizing spectral cues in a manner similar to NH listeners tend to be the ones who also show best word recognition performance.

IV. DISCUSSION

The objective of this study was to develop a test that could resolve differences in auditory spectral resolution for speech sound contrasts. The syllables /ba/ and /da/ were used as stimuli because they are contrasted primarily by spectral cues and are thus notoriously problematic for listeners with CIs. Formant transition cues and spectral tilt cues were systematically manipulated in modified natural speech, and the perceptual weighting of these cues was estimated using logistic regression at the group level, and empirical logit analysis at the individual level. Using results from listeners with NH as a model of optimal performance, it was shown that spectral degradation (whether deliberately induced via a vocoder or because of the use of CIs) reduced the weighting of formant transition cues, and increased the weighting of spectral tilt cues. Thus, the weighting of these cues can serve as an effective probe for spectral resolution. The trading of these cues in this study is consistent with the results of Alexander and Kluender (2009), which showed similar patterns for people with hearing impairment. It is reasonable to speculate that the limitations of sound frequency coding by CI processors results in gross reductions in reliability of formant peaks, leading listeners to shift weight to more reliable cues, as suggested by Toscano and McMurray (2010).

Previous phonetic cue-weighting studies with CI listeners suggest that limitations in the use of preferred acoustic cues can be complemented by increases in the use of secondary cues (Winn *et al.*, 2012; Moberly *et al.*, 2014), and the current study suggests that this phenomenon has implications for word recognition in general. CI listeners in this study who showed a cue-weighting pattern most similar to NH listeners in the unprocessed condition also tended to show best word recognition. However, this relationship was not especially strong; further work aimed explicitly at exploring this relationship could prove more fruitful in clarifying the connection between phonetic cue weighting and general success in speech perception. It is likely the case that intelligibility scores alone will not provide a comprehensive picture of such a relationship.

The modified phoneme recognition test used in this study was designed to avoid the monotony and unnaturalness of two-alternative forced choice tests, which comprise the majority of phonetic cue-weighting studies. Although responses for the /s, ʃ, r, l/ stimuli were not particularly informative for distinguishing among different conditions of resolution, they added the variety necessary to avoid abnormal fixation on a /b/-/d/ cue that might have arisen in a simple two-alternative test. These extra syllables also made the task seem easier, which is a noteworthy consideration for testing the CI population, who frequently undergo hours of experimentation and clinical evaluation, and who may feel a more personal attachment to their scores on auditory testing. As reported by many of the CI listeners in this study, the inclusion of a variety of syllables made the task more tolerable. It also likely cut down on unrealistically high accuracy that can be obtained when stimuli are repetitive and vary only slightly trial-to-trial.

The acoustic cue manipulation in this study was implemented over a range of frequencies that are known to be critical for speech intelligibility. Frequency-band importance functions provided by Kates (2013) suggest that energy between 800 and 4000 Hz contribute a disproportionately large amount to intelligibility. In this study, the formant cue was implemented within a narrow portion of this range, and the spectral tilt cue was implemented over a broader range centered on the peak of the aforementioned frequency-importance functions. Noting the history of the consonant place of articulation feature as the most susceptible to noise and degradation, the importance of this frequency band is sensible, because consonant place (as well as vowel place) is cued in large part by the second and third formant frequencies, which fall directly within the named frequency range.

Thus, when testing functional spectral resolution for speech sounds, it seems judicious to focus on this range, as it carries the most weight.

In this study, no attempt was made to quantify the spread of neural excitation in the cochlea stemming from activation of the electrodes that carried the critical spectral information relevant to the /ba/-/da/ contrast. One interpretation of the data obtained with CI users in the current study is that listeners who demonstrated poor performance (defined as little reliance on formant cues and greater reliance on spectral tilt) have relatively greater spread of neural excitation. This idea could be tested in future studies by measuring neural spread of excitation with electrical compound action potentials (Abbas *et al.*, 2004). Additionally, it is possible that poorer-performing CI listeners in this study would also demonstrate poorer performance on psychophysical tasks of electrode discrimination or pitch ranking.

One reported advantage of psychophysical tests over speech-based tests is that psychophysical tests are portable across languages (Jones *et al.*, 2013). Although the syllables used in this study were spoken by a speaker of North American English, the crucial /ba/-/da/ contrast is extremely common in the world's languages; it is contrastive in the 20 most popular languages spoken in the world, including every language with at least 50×10^6 first-language speakers (Lewis *et al.*, 2013), and is also contrastive among each of the 13 critical languages identified by the U.S. Department of State. Despite some cross-linguistic differences in the production of /d/,[2] it is likely that the current study could be portable across most languages.

There are some modest limitations of the test described in this paper. It is technically true that spectral resolution as classically defined can be independent of the behavior measured in this test. A listener could hypothetically perceive the spectral difference between each stimulus reliably, but simply choose to neglect such cues as a phonetic decision is made. Additionally, similar to spectral ripple tests (e.g., Won *et al.*, 2007), the test used in the current study does not provide frequency-specific information about a listener's hearing. The formant continuum in the current stimuli contained spectral peaks that spanned the frequency range from roughly 1000 to 2000 Hz, so it is likely that this frequency region is the one being assessed in the current study. However, the representation of the spectrum for CI listeners ultimately is determined by the engineering and signal processing of the speech processor. In view of that fact, the test merely measured the ability of CI listeners to make *use* of the information in the signal, regardless of how it is represented in their auditory systems.

Another limitation of the current study is the notable age difference between the NH and CI listener groups. This pattern is common in most CI research, mainly due to the availability of younger NH listeners and the lack of availability of younger CI listeners, who at this time are likely to have been congenitally deaf. It is reasonable to predict that, as implantation criteria continue to become less stringent, more young CI listeners will become available in the years to come. In the current experiment, most CI listeners had considerable experience with acoustic hearing before the onset of deafness, suggesting that they likely acquired typical cue-weighting strategy before the use of their implants. It remains unknown whether early-implanted children would show greater use of formant transitions for this consonant contrast, or if the limitations of the implant prevent mature-like cue-weighting strategies. Furthermore, it could be the case that, because of the lack of experience with well-defined formant cues, early-implanted children would show no preference for those cues, and instead rely mainly on those cues that are most reliably conveyed by the implant.

The relevance of spectral tilt as an acoustic cue has already been established in previous literature (cf. Blumstein *et al.*, 1982; Kiefte *et al.*, 2010; Alexander and Kluender 2008). The results of this study suggest that spectral tilt is an accessible cue for many listeners with CIs. In one sense, this is potentially beneficial information for the encoding of speech information by CI speech processors. However, the use of spectral tilt by CI listeners is likely limited by the well-documented limitations of the dynamic range in electric hearing (Zeng and Galvin, 1999). That is, global spectral level differences can only be represented within the range between threshold and maximum loudness transmittable by the implant processor. However, spectral tilt as a phonetic cue is potentially a parsimonious explanation for some examples of performance by CI listeners that exceeds the expectations set by a strictly frequency-based analysis of speech cues (Winn *et al.*, 2013b).

V. CONCLUSIONS

Functional spectral resolution for speech sounds can be measured as the perceptual weighting for formant transition cues for the /ba/-/da/ contrast, which represent a challenging spectral cue for a contrast known to be difficult for CI listeners. As spectral resolution is degraded, formant cues exert less influence on /ba/-/da/ perception, and listeners appear to compensate by making greater use of spectral tilt cues. Listeners with CIs demonstrate significantly less reliance on formant cues, and more reliance on spectral tilt compared to NH listeners. Given that these cues are complementary in natural speech, overall performance on word recognition might theoretically be unaffected by different cue-weighting strategies. However, CI listeners who showed cue-weighting most dissimilar to those of NH listeners also tended to have poorer word recognition.

The modified phoneme recognition task in this study tested specifically for the *mechanism* of phoneme identification rather than overall accuracy, and was thus able to resolve perceptual differences between CI listeners of different abilities who happen to show similar accuracy scores. Cue-weighting patterns can thus be exploited to evaluate the effectiveness of new CI signal processing strategies designed to improve spectral resolution. In view of the efforts by multiple CI manufacturers to address this problem (e.g., with a large number of electrodes, current steering, current focusing), such testing can play a vital role in the advancement of hearing technology.

ACKNOWLEDGMENTS

This work was supported by grants from the NIH-NIDCD: Grant No. R01 DC003083 (R.Y.L.), Grant No. R01 DC02932 (Edwards), and by a core grant to the Waisman Center from the NIH-NICHD (Grant No. P30 HD03352). M.B.W. is also supported by the NIH division of loan repayment. Figures in this paper were produced using ggplot2 (Wickham, 2009).

[1]Readers should note that in the cited version of the AngelSim software (V1.08.01), the carrier/synthesis filter slopes are implemented differently depending on whether the number of carrier channels (n) was equal to or less than the number of analysis channels (m). Selecting "6 dB/octave" for a vocoder where $n < m$ will yield 9 dB/ octave, and "18 dB/octave" for a vocoder where $n < m$ will yield roughly 50 dB/octave.

[2]There are some cross-linguistic differences in the realization of /d/, notably in terms of dentalization and exact timing of voicing, but the underlying /ba/-/da/ distinction in the spectral domain remains ostensibly the same.

Abbas, P., Hughes, M., Brown, C., Miller, C., and South H. (**2004**). "Channel interaction in cochlear implant users evaluated using the electrically evoked compound action potential," Audiol Neuro-Otol. **9**, 203–213.

Agresti, A. (**2002**). *Categorical Data Analysis*, 2nd ed. (Wiley, Hoboken, NJ), 168 pp.

Alexander, J., and Kluender, K. (**2008**). "Spectral tilt change in stop consonant perception," J. Acoust. Soc. Am. **123**, 386–396.

Alexander, J., and Kluender, K. (**2009**). "Spectral tilt change in stop consonant perception by listeners with hearing impairment," J. Speech Lang. Hear. Res. **52**, 653–670.

ANSI (**2010**). ANSI S3.6-2010, American National Standard Specification for Audiometers (American National Standards Institute, New York).

Aronoff, J., and Landsberger, D. (**2013**). "The development of a modified spectral ripple test," J. Acoust. Soc. Am. **134**, EL217–EL222.

Azadpour, M., and McKay, C. (**2012**). "A psychophysical method for measuring spatial resolution in cochlear implants," J. Assoc. Res. Otolaryngol. **13**, 145–157.

Barr, D. (**2008**) "Analyzing 'visual world' eye tracking data using multilevel logistic regression," J. Mem. Lang. **59**(4), 457–474.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (**2014**). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-6. http://CRAN.R-project.org/package=lme4 (Last viewed August 13, 2014).

Bierer, J. (**2007**). "Threshold and channel interaction in cochlear implant users: Evaluation of the tripolar electrode configuration," J. Acoust. Soc. Am. **121**, 1642–1653.

Bilger, R., and Wang, M. (**1976**). "Consonant confusions in patients with sensorineural hearing loss," J. Speech Hear. Res. **19**, 718–748.

Blumstein, S., Isaacs, E., and Mertus, J. (**1982**). "The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants," J. Acoust. Soc. Am. **72**, 43–50.

Boersma, P., and Weenink, D. (**2013**). "Praat: Doing phonetics by computer," [Computer program]. Version 5.3.56, retrieved September 15, 2013 from http://www.fon.hum.uva.nl/praat/ (Last viewed August 13, 2014).

Boëx, C., de Balthasar, C., Kós, M., and Pelizzone, M. (**2003**). "Electrical field interactions in different cochlear implant systems," J. Acoust. Soc. Am. **114**, 2049–2057.

Bonham, B., and Litvak, L. (**2008**). "Current focusing and steering: Modeling, physiology, and psychophysics," Hear. Res. **242**, 141–153.

Donaldson, G., and Nelson, D. (**2000**). "Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies," J. Acoust. Soc. Am. **107**, 1645–1658.

Dorman, M., and Loizou, P. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants," Am. J. Otolaryngol. **18**, 113–114.

Firszt, J., Koch, D., Downing, M., and Litvak, L. (**2007**). "Current steering creates additional pitch percepts in adult cochlear implant recipients," Otol. Neurotol. **28**, 629–636.

Francis, A., Kaganovich, N., and Driscoll-Huber, C. (**2008**). "Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English," J. Acoust. Soc. Am. **124**, 1234–1251.

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

Fu, Q.-J. (**2013**). "AngelSim: Cochlear implant and hearing loss simulator," [Computer program]. Version 1.07.01, retrieved April 15, 2013, available at http://www.tigerspeech.com/angelsim/angelsim_about.html (Last viewed August 13, 2014).

Goupell, M., Stoelb, C., Kan, A., and Litovsky, R. (**2013**). "Effect of mismatched place-of-stimulation on the salience of binaural cues in conditions that simulate bilateral cochlear-implant listening," J. Acoust. Soc. Am. **133**, 2272–2287.

Greene, B., Logan, J., and Pisoni, D. (**1986**). "Perception of synthetic speech produced automatically by rule: Intelligibility of eight text-to-speech systems," Behav. Res. Meth., Inst. Comp. **18**, 100–107.

Greenwood, D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Henry, B., Turner, C., and Behrens, A. (**2005**). "Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners," J. Acoust. Soc. Am. **118**, 1111–1121.

Hillenbrand, J., Clark, M., and Houde, R. (**2000**). "Some effects of duration on vowel recognition," J. Acoust. Soc. Am. **108**, 3013–3022.

Holt, L., and Lotto, A. (**2010**). "Speech perception as categorization," Attn., Percept. Psychophys. **72**, 1218–1227.

Jones, G., Won, J. H., Drennan, W., and Rubinstein, J. (**2013**). "Relationship between channel interaction and spectral-ripple discrimination in cochlear implant users," J. Acoust. Soc. Am. **133**, 425–433.

Kates, J. (**2013**). "Improved estimation of frequency importance functions," J. Acoust. Soc. Am. **134**, EL459–EL464.

Kiefte, M., Enright, T., and Marshall, L. (**2010**). "The role of formant amplitude in the perception of /i/ and /u/," J. Acoust. Soc. Am. **127**, 2611–2621.

Lewis, M. P., Simons, G., and Fennig, C. (**2013**). *Ethnologue: Languages of the World*, 17th ed. (SIL International, Dallas, TX). Online version: http://www.ethnologue.com (Last viewed August 13, 2014).

Lisker, L. (**1986**). "'Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees," Lang. Speech **29**, 3–11.

Litvak, L., Spahr, A., Saoji, A., and Fridman, G. Y. (**2007**). "Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners," J. Acoust. Soc. Am. **122**, 982–991.

McMurray, B., and Jongman, A. (**2011**). "What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations," Psych. Rev. **118**, 219–246.

Mens, L., and Berenstein, C. (**2005**). "Speech perception with mono- and quadrupolar electrode configurations: A crossover study," Otol. Neurotol. **26**, 957–964.

Mirman, D. (**2014**). *Growth Curve Analysis and Visualization Using R* (CRC Press, New York), pp. 109–112.

Moberly, A., Lowenstein, J., Tarr, E., Caldwell-Tarr, A., Welling, D., Shahin, A., and Nittrouer, S. (**2014**). "Do adults with cochlear implants rely on different acoustic cues for phoneme perception than adults with normal hearing?," J. Speech Lang. Hear. Res. **57**, 566–582.

Munson, B., Donaldson, G., Allen, S., Collison, E. A., and Nelson, D. A. (**2003**). "Patterns of phoneme misperceptions by individual with cochlear implants," J. Acoust. Soc. Am. **113**, 925–935.

Nelson, D., Van Tasell, D., Schroder, A., Soli, S., and Levine, S. (**1995**). "Electrode ranking of 'place-pitch' and speech recognition in electrical hearing," J. Acoust. Soc. Am. **98**, 1987–1999.

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**, 1777–1790.

Nusbaum, H., Pisoni, D., and Davis, C. (**1984**). "Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words," Research on Speech Perception Progress Report No. 10, Speech Research Laboratory, Indiana University, Bloomington, IN.

R Development Core Team (**2014**). "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria. [Computer software: version 3.1.0]. Available from http://www.R-project.org/ (Last viewed August 13, 2014).

Repp, B. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psych. Bull. **92**, 81–110.

Revoile, S., Holden-Pitt, L., and Pickett, J. (**1985**). "Perceptual cues to the voiced-voiceless distinction of final fricatives for listeners with impaired or with normal hearing," J. Acoust. Soc. Am. **77**, 1263–1265.
Saoji, A., Litvak., L., Spahr, A., and Eddins, D. (**2009**). "Spectral modulation detection and vowel and consonant identifications in cochlear implant listeners," J. Acoust. Soc. Am. **126**, 955–958.
Shannon, R., Fu, Q.-J., and Galvin, J. (**2004**). "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation," Acta Oto-Laryngol., Suppl. **552**, 50–54.
Shannon, R., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Singh, N., and Theunissen, F. (**2003**). "Modulation spectra of natural sounds and ethological theories of auditory processing," J. Acoust. Soc. Am. **114**, 3394–3411.
Srinivasan, A., Landsberger, D., and Shannon, R. (**2010**). "Current focusing sharpens local peaks of excitation in cochlear implant stimulation," Hear. Res. **270**, 89–100.
ter Keurs, M., Festen, J., and Plomp, R. (**1993**). "Effect of spectral envelope smearing on speech reception II," J. Acoust. Soc. Am. **93**, 1547–1552.
Toscano, J. C., and McMurray, B. (**2010**). "Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics," Cogn. Sci. **34**, 434–464.
Walsh, T., and Parker, F. (**1984**). "A review of the vocalic cues to [+- voice] in post-vocalic stops in English," J. Phonetics **12**, 207–218.
Wardrip-Fruin, C. (**1985**). "The effect of signal degradation on the status of cues to voicing in utterance-final stop consonants," J. Acoust. Soc. Am. **77**, 1907–1912.
Wickham, H. (**2009**). *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York), pp. 1–212.
Winn, M., Chatterjee, M., and Idsardi, W. (**2012**). "The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing," J. Acoust. Soc. Am. **131**, 1465–1479.
Winn, M., Chatterjee, M., and Idsardi, W. (**2013a**). "Effects of masking noise and low-pass filtering," J. Speech Lang. Hear. Res. **56**, 1097–1107.
Winn, M., Rhone, A., Chatterjee, M., and Idsardi, W. (**2013b**). "Auditory and visual context effects in phonetic perception by normal-hearing listeners and listeners with cochlear implants," Frontiers Psych. Auditory Cogn. Neurosci. **4**, 1–13.
Won, J., Drennan, W., and Rubinstein, J. (**2007**). "Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users," J. Assoc. Res. Otolaryngol. **8**, 384–392.
Xu, L., Thompson, C., and Pfingst, B. (**2005**). "Relative contributions of spectral and temporal cues for phoneme recognition," J. Acoust. Soc. Am. **117**, 3255–3267.
Zeng, F.-G., and Galvin, J. (**1999**). "Amplitude compression and phoneme recognition in cochlear implant listeners," Ear Hear. **20**, 60–74.
Zwolan, T., Collins, L., and Wakefield, G. (**1997**). "Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects," J. Acoust. Soc. Am. **102**, 3673–3685.