Individual Variability in Recalibrating to Spectrally Shifted Speech: Implications for Cochlear Implants

Michael L. Smith[1] and Matthew B. Winn[2]

Objectives: Cochlear implant (CI) recipients are at a severe disadvantage compared with normal-hearing listeners in distinguishing consonants that differ by place of articulation because the key relevant spectral differences are degraded by the implant. One component of that degradation is the upward shifting of spectral energy that occurs with a shallow insertion depth of a CI. The present study aimed to systematically measure the effects of spectral shifting on word recognition and phoneme categorization by specifically controlling the amount of shifting and using stimuli whose identification specifically depends on perceiving frequency cues. We hypothesized that listeners would be biased toward perceiving phonemes that contain higher-frequency components because of the upward frequency shift and that intelligibility would decrease as spectral shifting increased.

Design: Normal-hearing listeners (n = 15) heard sine wave-vocoded speech with simulated upward frequency shifts of 0, 2, 4, and 6 mm of cochlear space to simulate shallow CI insertion depth. Stimuli included monosyllabic words and /b/-/d/ and /ʃ/-/s/ continua that varied systematically by formant frequency transitions or frication noise spectral peaks, respectively. Recalibration to spectral shifting was operationally defined as shifting perceptual acoustic-phonetic mapping commensurate with the spectral shift. In other words, adjusting frequency expectations for both phonemes upward so that there is still a perceptual distinction, rather than hearing all upward-shifted phonemes as the higher-frequency member of the pair.

Results: For moderate amounts of spectral shifting, group data suggested a general "halfway" recalibration to spectral shifting, but individual data suggested a notably different conclusion: half of the listeners were able to recalibrate fully, while the other halves of the listeners were utterly unable to categorize shifted speech with any reliability. There were no participants who demonstrated a pattern intermediate to these two extremes. Intelligibility of words decreased with greater amounts of spectral shifting, also showing loose clusters of better- and poorer-performing listeners. Phonetic analysis of word errors revealed certain cues were more susceptible to being compromised due to a frequency shift (place and manner of articulation), while voicing was robust to spectral shifting.

Conclusions: Shifting the frequency spectrum of speech has systematic effects that are in line with known properties of speech acoustics, but the ensuing difficulties cannot be predicted based on tonotopic mismatch alone. Difficulties are subject to substantial individual differences in the capacity to adjust acoustic-phonetic mapping. These results help to explain why speech recognition in CI listeners cannot be fully predicted by peripheral factors like electrode placement and spectral resolution; even among listeners with functionally equivalent auditory input, there is an additional factor of simply being able or unable to flexibly adjust acoustic-phonetic mapping. This individual variability could motivate precise treatment approaches guided by an individual's relative reliance on wideband frequency representation (even if it is mismatched) or limited frequency coverage whose tonotopy is preserved.

[1]Department of Speech & Hearing Sciences, University of Washington, Seattle, Washington, USA; and [2]Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis, Minnesota, USA.

Key words: Cochlear implants, Individual differences, Perceptual adaptation, Phonetic perception, Recalibration, Spectral shift, Speech perception, Vocoded speech.

(Ear & Hearing 2021;XX;00–00)

INTRODUCTION

Cochlear implants (CIs) provide access to sound for individuals with severe to profound hearing loss and are considered to be one of the more successful neural prostheses (Wilson & Dorman 2008). However, outcomes have been shown to be extremely variable, with speech understanding ranging from 0 to 100% intelligibility (Blamey et al. 2013). High variability in outcomes of CI listeners has been the focal point of research for several years, and the underlying cause of this variability has been shown to be multi-faceted (Lazard et al. 2012). For example, there is variability among etiology of deafness, health of the cochlea, duration of deafness, age of onset of hearing impairment, and the biophysical interactions between the electrodes and the neurons (Bierer et al. 2011; Bierer & Litvak 2016). Spectral resolution is notoriously poor in CIs because of the limited number of electrodes and spread of electrical activity throughout the cochlea. Holden et al. (2013) found that placement and insertion depth of the electrode array were one of the most significant factors that affected patient outcomes. Incomplete insertion of the CI into the cochlea results in upward shifting of the frequency spectrum. Landsberger et al. (2015) estimated that the mean apical place of stimulation would correspond to 323 to 740 Hz for various CI devices. The exact amount of frequency mismatch resulting from these stimulations depends partly on the analysis bands but can be roughly estimated to be about 0.5 to 0.9 octaves at the apex, with smaller mismatches toward the base. The problem resulting from shallow CI insertion depth is that the representation of frequencies will be tonotopically mismatched. Unknown is whether some listeners are more susceptible to struggling with this mismatch, in addition to poor frequency resolution of the implant.

Poor representation of spectral information in CIs has consequences for perceiving speech, as many speech sounds (e.g., /b/-/d/, /ʃ/-/s/, vowels) are distinguishable by changes in the spectral domain. Changes in consonant place of articulation (PoA) result in changes in the frequency spectrum, and perception of those spectral differences are essential for a listener's ability to identify which sound was spoken. Due to poor frequency representation of the implant, and thus poor spectral resolution, sensitivity to changes in PoA is typically the most difficult speech feature for CI listeners to perceive (Munson & Nelson 2005; Munson et al. 2003; but see Rødvik et al. 2019 for extra nuance). In addition, normal-hearing (NH) listeners have also shown to have difficulty perceiving differences of PoA when listening to speech that is spectrally degraded (Xu et al. 2005; Zhou et al. 2010; Winn & Litovsky 2015). Consonant PoA will therefore be the focus of the current investigation.

<zdoi; 10.1097/AUD.0000000000001043>

We expect that a postlingually deafened recipient of a CI will need to undergo phonetic "recalibration," which refers to the need to adjust acoustic-phonetic frequency boundaries commensurate with the shifting of the spectrum as delivered by the CI. For example, if there is an important frequency boundary at 2000 Hz but the spectrum has shifted up by an octave, the listener ought to correspondingly shift the perceptual boundary up to 4000 Hz to maintain the proper distinctions in the transformed input. Conversely, the lack of calibration would mean that the listener maintains rigid frequency boundaries for phonetic contrasts despite wholesale changes in the spectrum. Given these two possibilities recalibration to a frequency mismatch is the optimal strategy and is the focus of the present study.

Due to the small scale and the nonlinear tonotopic spacing of the cochlea, a shift of just 1 or 2 mm of array insertion can result in a frequency shift of hundreds or thousands of Hz (Greenwood 1990). Unsurprisingly, shallow electrode insertion depth results in significant difficulty in the identification of consonants (Fu et al. 2002; Li & Fu 2010), vowels (Fu & Shannon 1999), words (Fu et al. 2002), and sentences (Dorman et al. 1997). Still, some very basic questions about the consequence of spectral shifting remain unknown, because the experimental stimuli were not necessarily designed to reveal whether a listener properly adjusted acoustic-phonetic mapping. There are rare examples of this priority represented in the literature, but some are constrained to multidimensional frequency space that is difficult to interpret (cf. Harnsberger et al. 2001) or limited to simulation studies to exert rigorous control over the frequency regions that are affected (DiNino et al. 2016). It is important to note that it has been observed that spectral shifting results in unique challenges separate from spectral degradation. For example, Fu and Shannon (1999) showed that vowel recognition deficits resulting from spectral shifting were not alleviated by improvements in spectral resolution.

It is clear from everyday experience that listeners can accommodate some amount of spectral shifting when identifying speech, because differences in vocal tract size produce systematic proportional differences in voice resonant frequencies (i.e., formants). As individuals with typical hearing do not encounter substantial difficulty with talkers of different sizes (and hence with shifted patterns of spectral peaks in their voices), we can expect some robustness of perception across a moderate amount of artificial spectral shifting. However, the amount of frequency shift resulting from a shallow CI insertion depth far exceeds the amount of spectral shifting that occurs naturally from differences in speaker vocal tract size among typical adults and children (Simpson 2009; Story et al. 2018). Differences of the second formant for the vowel /i/ produced by women and men are among the largest gender-related frequency differences and yet only reflect an approximate change of 435 Hz or 0.25 octaves across gender (Hillenbrand et al. 1995). The corresponding shift in frequency energy from the male to female talker would correspond to a shift of just over 1 mm of cochlear space according to Greenwood's (1990) function. Conversely, spectral shifting in a CI is estimated to be in the range of 3.3 to 6 mm at the apical end, when comparing the frequencies stimulated by the most apical angular insertion depth (Landsberger et al. 2015) and standard frequency allocation at the most apical electrode. Therefore, listening with a CI demands recalibration to a degree not encountered in everyday speech communication when simply interacting with a variety of talkers.

Summary and Hypothesis

Prior studies have shown spectral shifting is an expected consequence of using a CI and that it has some degree of impact on speech intelligibility (Dorman et al. 1997; Fu & Shannon 1999; Fu et al. 2002; Li & Fu 2010). Based on the large degree of frequency shift, adjustment to CI-like frequency mismatches demand perceptual flexibility far exceeding that which is needed for encountering an everyday variety of talkers. However, it remains unclear whether individuals vary in their ability to adjust their perception in response to frequency shifting. A more detailed examination of this ability is needed, which could be aided by a specific focus on the perception of consonants that differ by frequency spectra—which conveniently happen to be the consonant contrasts that are most problematic for CI listeners.

We hypothesize that (1) perception of the /ʃ/-/s/ and /b/-/d/ phoneme contrasts will become biased specifically toward /s/ and /d/, respectively, because they contain relatively higher-frequency components than their counterparts and (2) based on some evidence from previous studies showing individual variability in speech understanding to spectrally shifted stimuli some participants will be better at recalibrating to the spectrally shifted stimuli than others, and this will emerge as relatively better performance in both the categorization of phonemic continua (in terms of reduced bias toward /s/ and /d/) and word intelligibility. We operationally defined recalibration to a spectral shift as the adjustment of acoustic-phonetic mapping commensurate with the shifting of the frequency spectrum. Listeners who recalibrate are able to maintain perceptual differentiation between phonemes even after the entire continuum has been spectrally shifted. Conversely, a listener who fails to recalibrate is hypothesized to show categorization bias toward perceiving the phoneme with high-frequency components (e.g., /s/ instead of /ʃ/) when the spectrum is shifted upward.

MATERIALS AND METHODS

Participants

The study included 15 NH subjects (mean age = 27 years; range = 20 to 47 years) who were screened to confirm pure-tone thresholds of ≤20 dB HL at octave frequencies between 0.25 and 8 kHz in both ears (American National Standards Institute Accredited Standards Committee S3, Bioacoustics 2004). Extended high-frequency hearing (>8 kHz) was not tested (elaborated further in the Discussion section). All experimental protocols were approved by the Institutional Review Board of the University of Washington, and all listeners provided informed written consent before participation. All listeners were compensated for their participation. They all spoke North American English as their native language and did not self-report any language-learning or cognitive deficits.

Stimuli

Speech stimuli consisted of modified natural speech tokens that were spoken by a native adult male speaker of American English. There were three groups of sounds—a /bɑ/-/dɑ/ contrast, a /ʃɑ/-/sɑ/ contrast, and a /rɑ/-/lɑ/ contrast, each described later, as well as monosyllabic words. Testing was completed in a sound-attenuated booth with stimuli delivered via a Tannoy Reveal 402 loudspeaker (frequency response 56 to 48,000 Hz ± 3 dB SPL) at 0 degrees azimuth at 65 dBA.

Monosyllabic Words

Open-set word recognition was tested using monosyllabic words (e.g., "boat," "dime," "take," "run") in each of the vocoder conditions described later. The words were drawn from the Maryland consonant-nucleus-consonant (CNC) clinical corpus (Peterson & Lehiste 1962), which was designed to contain words that are familiar to most listeners and yet difficult enough to differentiate among various degrees of hearing impairment (e.g., mild/severe/profound). The words were presented in isolation in quiet, with no carrier phrase.

/bɑ/-/dɑ/ Continuum

The /b/-/d/ continuum was divided into eight steps (i.e., differences that were equally spaced on a psychoacoustic [Bark] frequency continuum) and featured manipulation of formant frequency transitions at the onset of the syllable to cue the phonetic contrast. The second formant changed a total of 0.8 octaves, from roughly 1000 Hz (at the /b/-endpoint) to 1800 Hz (at the /d/ endpoint), with smaller correlated changes in F3 as well. The stimuli were a subset of those described and illustrated by Winn and Litovsky (2015), excluding any variations in spectral tilt that were examined in that study; greater detail on the creation of these stimuli can be found in their paper. In short, naturally uttered syllables were manipulated using the linear predictive coding decomposition method in Praat (Boersma & Weenink 2013), so that formant trajectories intermediate to /bɑ/ and /dɑ/ could be superimposed on the residual voice source from the original utterance and refiltered to produce natural-sounding speech with parametrically controlled formant transitions. Consonant release bursts were filtered by the onset of the formant contours and thus complimented the formant transitions for a given continuum step.

/ʃɑ/-/sɑ/ Continuum

The /ʃ/-/s/ continuum was used to have a phoneme contrast whose information-bearing frequency content was spread across a wider frequency range (between 2000 and 8000 Hz; two octaves), as frication energy is less compact than formant frequencies in /b/ and /d/. There were seven steps in the /ʃ/-/s/ continuum, where the endpoints were natural /ʃ/ and /s/ sounds. Consistent with the naturally produced signals, the fricative spectrum peaks varied in terms of relative amplitude of spectrum peaks within the fricative noise. In general, there is a lower-frequency peak (around 2300 Hz) that has relatively greater amplitude in the /ʃ/ token and a higher-frequency peak (around 5500 Hz) that has relatively greater amplitude in the /s/ token. Intermediate steps were created by gradual blending of these signals mixed at different proportions (e.g., 84% /s/ + 16% /ʃ/, 68% /s/ + 32% /ʃ/, etc.). Each member of the continuum was pre-appended onto the vowel from a natural production of /sɑ/, chosen because it is perceptually compatible with either fricative, whereas the vowel following /ʃɑ/ would sound slightly unnatural when preceded by /s/ (i.e., it would sound like "sya" because of the palatal tongue position at the /ʃ/-/ɑ/ boundary).

/rɑ/-/lɑ/ Sounds

The two remaining sounds were /rɑ/ and /lɑ/, which were unaltered natural recordings of these syllables. Consistent with the approach used by Winn and Litovsky (2015), these sounds were included simply to introduce more variety to make the task less monotonous and to avoid unnaturally focused attention solely on the cues for the previous contrasts. Their inclusion also conferred the additional benefit of making the task appear easier and more engaging.

Vocoder Parameters

Although a vocoder simulation does not perfectly replicate the hearing of a CI user, it can provide researchers with the ability to systematically manipulate various aspects of an auditory signal that would not be under experimenter control in a real sample of CI listeners, including spectral resolution and spectral shifting. Sine wave vocoding was done similarly to what has been reported in earlier studies (Dorman et al. 1998; Li et al. 2009; Li & Fu 2010). In short, the frequency spectrum was divided into a discrete number of analysis channels, and the amplitude envelope of each channel was imposed upon a sine wave whose frequency corresponded to the center of each analysis band. Absolute lower- and upper-frequency boundaries were set to 100 and 8000 Hz, respectively, to capture the speech frequency range. A total number of 16 frequency channels were used, which has been shown to allow near-perfect intelligibility performance in quiet (Dorman et al. 1998). We wish to note, however, that high intelligibility does not necessarily mean perfect perception of all acoustic-phonetic cues or robustness to spectral shifting. Fewer frequency channels were not used (i.e., less spectral resolution) because the similarity to CI performance is mainly restricted to word intelligibility and not necessarily the perception of individual phonetic cues.

A sine wave carrier was used instead of a noise carrier to more faithfully preserve temporal envelope modulations of the speech signal, with envelopes of each channel low-pass filtered at 600 Hz to preserve periodicity and high-speed amplitude envelope changes. Temporal envelopes of a sine wave carrier lack the temporal distortions that result from noise carriers that would contain envelope fluctuations inherent to any filtered band of Gaussian noise (Kohlrausch et al. 1997; Stone et al. 2008; Oxenham & Kreft 2014).

Spectral Shifting

To simulate different amounts of electrode insertion depth, the vocoder simulations incorporated upward shifting of carrier channel frequencies away from the center of the analysis channels. The Greenwood (1990) frequency-to-place equation was used to determine the frequencies that corresponded to shifts of 2, 4, and 6 mm of cochlear space. Figure 1 displays the center frequencies of each channel and how they were shifted in each condition. In this study, all frequencies were shifted by a uniform cochlear distance, although frequency alignment in a real CI is complicated by numerous other factors. The highest frequency carrier channel in the 6 mm condition was 16,581 Hz, which extends roughly one octave beyond the audiometric testing thresholds used to screen for hearing loss in the present study (further comments in the discussion). However, it is not necessary to perceive the highest-frequency channel to recognize the differences between the two endpoint stimuli, as the phonetic differences were also contrastive within lower channels.

Spectrograms of /ʃɑ/-/sɑ/ and /bɑ/-/dɑ/ for the 0 and 6 mm conditions are shown in Figure 2 and Figure 3, respectively. The information-bearing frequency region for each contrast is highlighted by a gray rectangle. Notice how the lower end of this

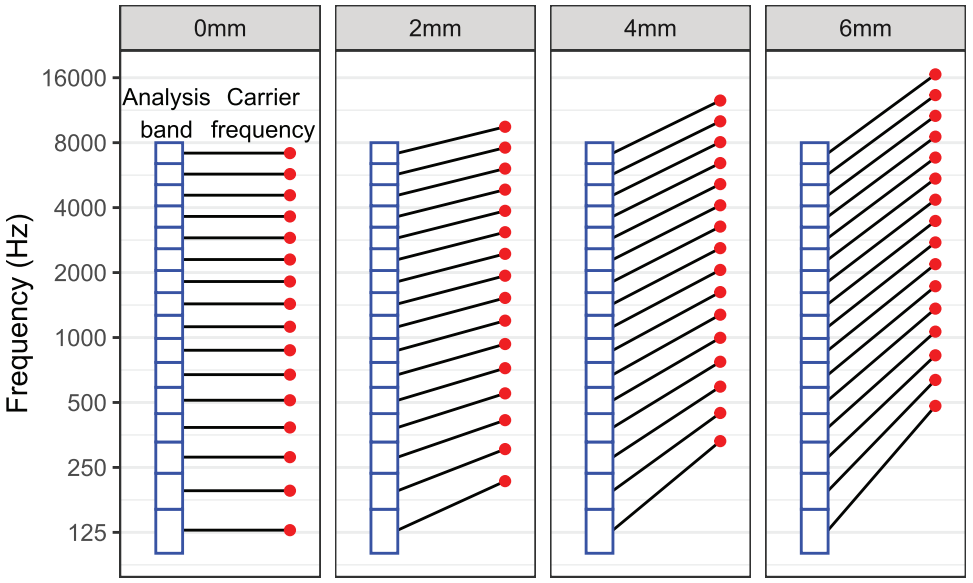

Fig. 1. Center frequencies for each channel and how they were shifted in each condition. The vertical span of the boxes represents the analysis band. The points extending from each box represent the carrier frequency of the respective synthesis band. Each panel represents a unique condition, with the 0 mm condition representing a vocoded condition without spectral shifting. The subsequent panels represent shifts of 2, 4, and 6 mm of cochlear space, respectively.

region for a 6 mm shifted /ʃɑ/ or /bɑ/ now overlaps with the upper end of the unshifted /sɑ/ or /dɑ/ frequency range. Thus, listeners will need to recalibrate to this frequency mismatch to perceive two distinct phonetic categories of perception rather than classifying all of the /ʃɑ/-/sɑ/ sounds as /sɑ/ or all of the /bɑ/-/dɑ/ sounds as /dɑ/.

Procedure

Listeners completed both the phoneme categorization task and the word identification task in five conditions: normal (not vocoded), vocoded with 0, 2, 4, and 6 mm shifts. Stimuli were blocked by the degree of spectral shifting and the ordering of blocks was randomized.

Monosyllabic words were tested by using 50 words played in each vocoder condition, with each block selecting words randomly from the 400-word CNC corpus and a separate word list produced for each vocoder condition to avoid learning effects for memorized lists. Listeners heard one word at a time and responded by typing the word into a computer interface. Blocks of CNCs were randomly interspersed between testing blocks of phoneme categorization.

Phoneme categorization was tested using a single-interval six-alternative forced-choice task (hear one syllable at a time, and label it using the choices /sɑ/, /ʃɑ/, /bɑ/, /dɑ/, /rɑ/, /lɑ/), and all six choices were on the computer screen at one time. Listeners responded by selecting their answer via mouse-click. There were three blocks presented for each condition, and each

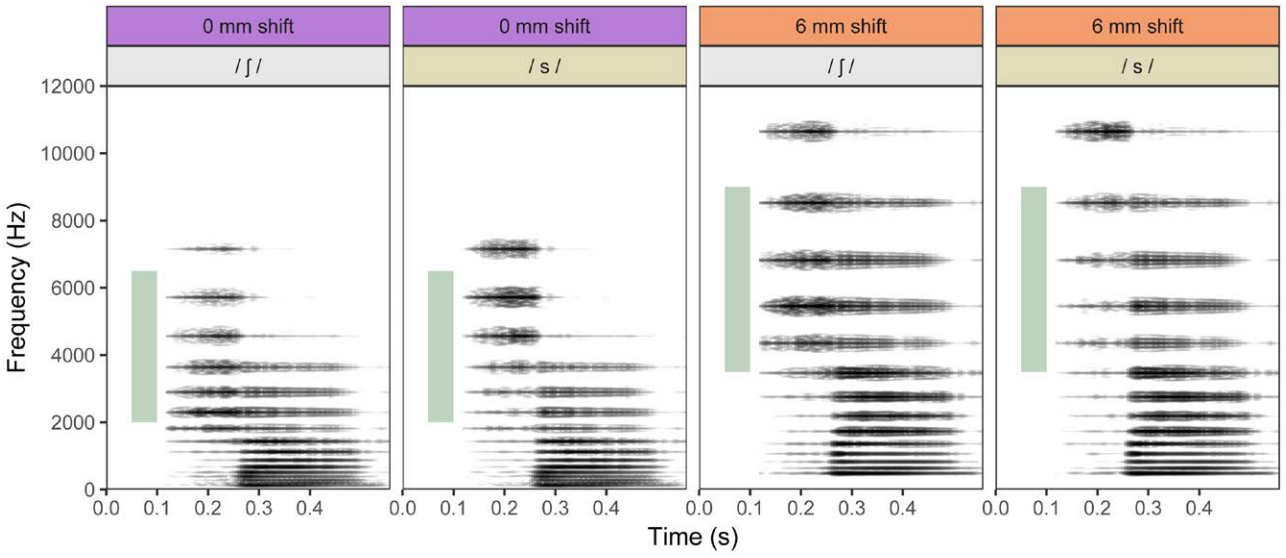

Fig. 2. Spectrograms for the /ʃɑ/-/sɑ/ contrast, showing continuum endpoints that reflect typical productions of these phonemes. Each panel represents an entire syllable in either the 0 or 6 mm condition. The gray rectangle in the spectrogram represents the frequency range that contains the contrastive frequency cue for the pair of phonemes. Note how the upper edge of the rectangle for the 0 mm condition overlaps with the lower edge of the rectangle for the 6 mm condition.

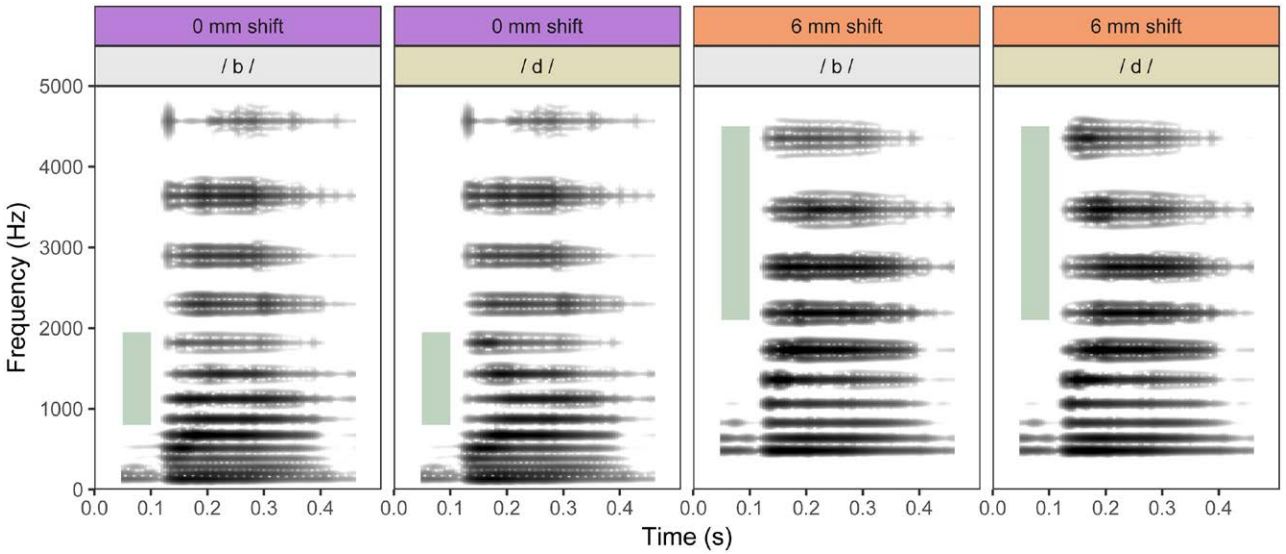

Fig. 3. Spectrograms for the /bɑ/-/dɑ/ contrast, showing continuum endpoints that reflect typical productions of these phonemes. Each panel represents an entire syllable in either the 0 or 6 mm condition. The gray rectangle in the spectrogram represents the frequency range that contains the contrastive frequency cue for the pair of phonemes. Note how the upper edge of the rectangle for the 0 mm condition is close to the lower edge of the rectangle for the 6 mm condition.

block contained three trials for every unique step of both the /bɑ/-/dɑ/ and /ʃɑ/-/sɑ/ continua, respectively, yielding nine observations for each unique continuum step. Each block also contained 8 trials each for /lɑ/ and /rɑ/, for a total of 61 trials per block, 183 trials per condition, for a grand total of 732 trials. The /bɑ/-/dɑ/ and /ʃɑ/-/sɑ/ contrasts were slightly overrepresented within each block (39% and 35%, respectively, rather than 33%), so more data were collected on the specific measures of interest. The /bɑ/-/dɑ/ and /ʃɑ/-/sɑ/ stimuli were members of gradual continua, but participants were not made aware of these manipulations, nor were they aware of the insignificance of the /rɑ/-/lɑ/ trials. Listeners could choose to repeat a stimulus presentation when needed but were encouraged to answer based on their first impression and only repeat when a stimulus was missed (e.g., because of a cough). Only 2.7% of trials in the unprocessed condition contained a stimulus repeat, with only 2.5%, 3.3%, 5.6%, and 6.3% of trials repeated in the 0, 2, 4, and 6 mm conditions, respectively.

Each participant began the experiment with a short practice block for the nonvocoded and the 2 mm shifted conditions. This was done to familiarize listeners with the experiment interface, stimuli, and overall procedure. After the practice blocks, each listener's first block was always one of the three normal (unprocessed) condition blocks, with the rest of the subsequent blocks randomized across conditions and task (phonemes or CNC words). Total testing time was approximately 90 minutes including breaks when needed.

Analysis

Monosyllabic Words • Individual CNC words were scored as either correct or incorrect. Alternate spellings and homophones (e.g., sore-soar) were manually scored to count as correct in the analysis. An analysis of varying phonetic features was conducted on each consonant in each CNC word to better understand the impact of spectral shifting on recovering specific features like voicing, manner of articulation (MoA), and PoA. Each feature was tracked independently, so that there could be more than one feature error on each consonant. For example, a misperception of /b/ as /k/ would be counted as both a voicing error and a PoA error. This analysis was conducted separately for consonants in the word-onset or word-final position.

Binomial Model of Phoneme Categorization • For each continuum, listeners' identification responses were modeled using generalized linear (logistic) mixed-effects model (GLMM) using the lme4 package (1.1-21; Bates et al. 2015) in the R software interface (R Core Team 2016). There were separate models for the /b/-/d/ and the /s/-/ʃ/ continua, with each model estimating the log odds of a listener perceiving the higher-frequency member of each pair (/d/ or /s/) as a function of the continuum step and condition (shift), and their fully crossed interactions. Each fixed effect was also declared as a random effect to account for dependence across repeated measures and to reduce type 1 error rate.

Four-Parameter Model of Phoneme Categorization • Psychometric functions did not always asymptote at floor (0) and ceiling (1) in the vocoder conditions, which complicates interpretation of the binomial modeling described earlier. Therefore, psychometric functions of perceived PoA were additionally modeled as four-parameter sigmoidal functions that allowed the asymptotes to deviate from floor and ceiling. A nonlinear least squares curve fitting procedure was used, incorporating the Levenberg-Marquardt convergence algorithm available in the "minipack.lm" package in R. The form of the function was as follows:

$$\text{Alveolar} = \left[\text{Range} / 1 + e^{\left(-(\text{step}+\times 50)*\text{slope}\right)}\right] + \text{Floor}$$

"Alveolar" refers to the perceived PoA by the listener as the higher-frequency member of the pair (e.g., /dɑ/ instead of /bɑ/ or /sɑ/ instead of /ʃɑ/). "Floor" refers to the lower asymptote of the psychometric function, which could be as low as 0. "Range" refers to the difference between the floor and the upper asymptote of the function, which could potentially span up to 1 if the floor is 0. "x50" refers to the x axis position (continuum step) along the function at which the estimated value is halfway

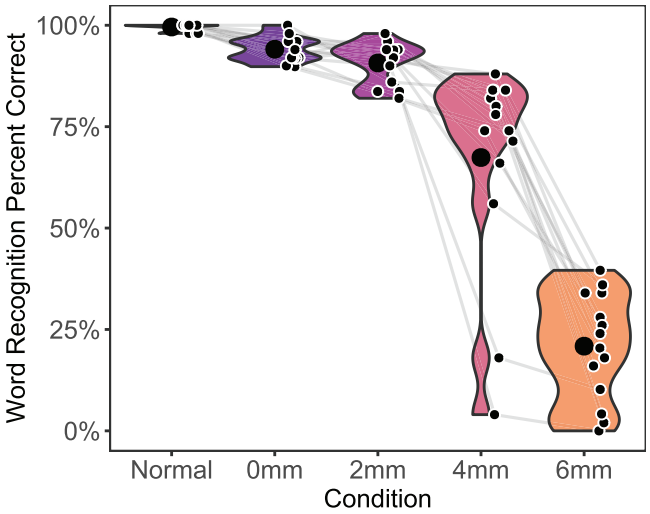

Fig. 4. Word intelligibility performance across listening conditions. The large black dot represents the group mean, with individual data points represented by the small black dots. Lines connect data for individuals across conditions. Width of the violin shape underneath the dots reflects the density distribution of the individual data.

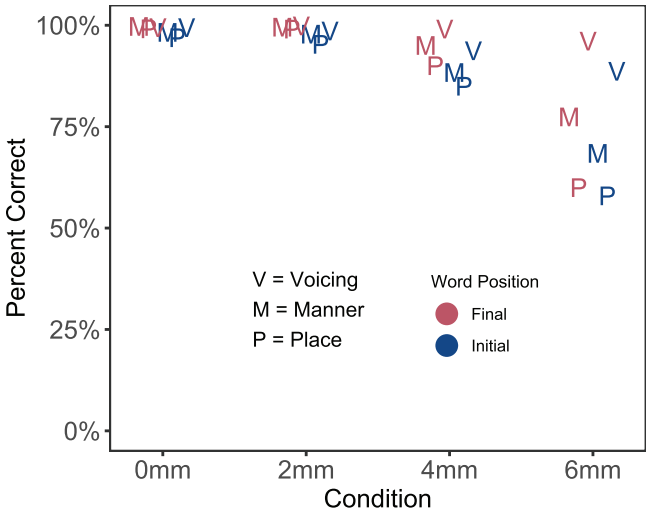

Fig. 5. Percent correct of each consonant feature for both word-initial (dark blue) and word-final (light red) position for the features of manner (M), place (P), and voicing (V) in each vocoded listening condition. For example, in the 6 mm condition, the manner of articulation at word-final position was perceived correctly 77% of the time.

between the lower and upper asymptote. "Step" refers to the continuum step that was presented. All four terms of the model were free to vary across individuals and modeled as interactions with the vocoder condition. Allowing individually free variable estimation provided much better fits to the data, as expected based on visual inspection of the figures. The added value of this four-parameter approach is the explicit estimation of individuals' floor and ceiling parameters, which ultimately became a centerpiece of the interpretation of this study.

RESULTS

Word Identification Task

Overall Word Intelligibility • Word intelligibility scores decreased as a function of spectral shifting, shown in Figure 4. Average group performance (large black dot) was virtually at 100% in the normal condition, 94% for the 0 mm shift condition, 91% with 2 mm of shifting, 67% in the 4 mm condition, and 21% in the 6 mm condition. Individual performance showed a wider range of variability in the 4 and 6 mm conditions.

Phonetic Features • Analysis of the perception of phonetic features (Fig. 5) revealed that the decline in performance across listening conditions was not uniform across the features. PoA was the most fragile feature in all vocoder conditions, consistent with earlier literature (e.g., Miller & Nicely 1955; Xu et al. 2005). Perception of voicing in word-final consonants was generally shielded from the detrimental effects of spectral shifting. Notably, among all of the shifted conditions, /ʃ/ was perceived as /s/ 13% of the time (12\96 opportunities), while /s/ was misperceived as /ʃ/ only once out of 240 opportunities, and never in word-onset position. In those shifted conditions, /ʃ/ had 81% overall accuracy while /s/ had 91% accuracy. Accuracy for /b/ was 74% overall. Contrary to our expectations, misperception of /b/ as /d/ never occurred in the word-onset position but was the most frequent error in the word-final position, with a roughly 20% error rate. Across all shifted conditions, /d/ had 79% overall accuracy; it was perceived as /b/ and /g/ with rates of 4% and 8%, respectively. Nasal sounds—containing predominantly low-frequency energy, were perceived with 74% accuracy in the shifted conditions. The most common error on nasal sounds was misperceiving /m/ as /n/ (at a 19% rate), which is consistent with the pattern that was expected for the /b/-/d/ sounds (which are simply oral variants of these nasal sounds). Conversely, /n/ was rarely misperceived as /m/, with an error rate of less than 2%. These error patterns are generally consistent with the hypothesis that misperceptions of shifted phonemes will gravitate toward similar phonemes with higher-frequency spectral components.

Table 1 shows the GLMM results for the phonetic feature analysis. In this table, we are modeling the change in log odds of correct perception of a given phonetic feature (manner, place, and voicing), given specific changes in stimulus parameters. The default parameters of the model were 0 mm shifting, word-offset position, and MoA. The outcome for any other feature, word position or listening condition represents a deviation from these defaults. To explain how to interpret the model coefficients, we will first walk through some simple and complex examples. Line β4 in the table represents a single parameter deviation from the default, from 0 to 6 mm, keeping the defaults of word-offset and MoA. Therefore, line β4 does not reflect an "overall" effect of 6 mm shifting; it reflects how perception of word-offset MoA differs between the 0 and 6 mm conditions. At word-offset position, comparison between perception of manner of articulation to place of articulation involves line β9. When analyzing perception of MoA in the 0 mm (unshifted) stimuli, comparison of performance at word-onset to word-offset position involves line β5.

The estimate of performance for word-onset PoA in the 4 mm shifted condition involves summing the intercept (β1; default parameters), the 4 mm main effect (β3), the onset position main effect (β5), the PoA main effect (β9), the interaction between 4 mm and onset position (β7), the interaction between 4 mm and PoA (β11), the interaction between onset position and PoA (β13), and the interaction between 4 mm and PoA and onset position (β15). Together, these coefficients sum to 1.95 log odds of correct perception for word-onset PoA in the

TABLE 1. Generalized linear mixed-effects model describing performance in perceiving phonetic features as a function of feature type, word position, and vocoder condition

Term	Feat	Pos	mm	Term	Est	Total Coef	%	SE	*F*	*p*
β1	MoA	Offset	0	Intercept (feature: MoA)	5.88	5.88	99.7	0.73	8.10	< .001
β2			2	2 mm	–0.14	5.74	99.7	0.98	–0.14	.890
β3			4	4 mm	–2.23	3.64	97.5	0.84	–2.65	.007
β4			6	6 mm	–4.62	1.26	77.9	0.79	–5.88	< .001
β5		Onset	0	Onset position	–1.72	4.16	98.5	0.80	–2.16	.031
β6			2	2 mm: onset	–0.01	4.01	98.2	1.04	–0.01	.989
β7			4	4 mm: onset	0.39	2.31	91.0	0.86	0.46	.648
β8			6	6 mm: onset	1.21	0.75	67.9	0.82	1.48	.138
β9	PoA	Offset	0	Feature: PoA	–1.09	4.79	99.2	0.84	–1.30	.194
β10			2	2 mm: PoA	0.32	4.97	99.3	1.12	0.28	.776
β11			4	4 mm: PoA	0.00	2.55	92.8	0.88	0.00	.999
β12			6	6 mm: PoA	0.15	0.32	57.9	0.86	0.17	.862
β13		Onset	0	Onset position: PoA	0.53	3.59	97.3	0.92	0.57	.570
β14			2	2 mm: PoA: onset	–0.62	3.14	95.8	1.24	–0.50	.615
β15			4	4 mm: PoA: onset	0.19	1.94	87.4	0.97	0.20	.841
β16			6	6 mm: PoA: onset	–0.10	0.09	52.2	0.94	–0.10	.916
β17	V	Offset	0	Feature: voicing	–0.48	5.40	99.5	0.92	–0.52	.601
β18			2	2 mm: voicing	1.92	7.19	99.9	1.57	1.22	.221
β19			4	4 mm: voicing	2.08	5.24	99.5	1.04	2.00	.045
β20			6	6 mm: voicing	2.81	3.59	97.3	0.97	2.91	.003
β21		Onset	0	Onset position: voicing	1.84	5.52	99.6	1.08	1.70	.088
β22			2	2 mm: voicing: onset	–2.62	4.68	99.1	1.77	–1.48	.138
β23			4	4 mm: voicing: onset	–2.62	3.13	95.8	1.22	–2.15	.031
β24			6	6 mm: voicing: onset	–2.77	2.16	89.6	1.15	–2.40	.016

Est, beta estimate; Feat, phonetic feature; MoA, manner of articulation; PoA, place of articulation; Pos, word position (onset or offset); Total Coef, linear sum of all interacting coefficients for the term in the model; V, voicing.

4 mm shifted condition. The equivalent proportional response that would correspond to these log odds is calculated using the inverse logit function 1/(1 + exp(-log odds)). An argument of 1.95 entered into this function yields 87.5%, which is exactly the performance achieved for this condition (line 15, % column). The GLMM thus represents the exercise of decomposing these performance scores into unique contributions of each of the phonetic parameters and listening conditions.

Only a few specific effects/interactions reached statistical detection in the full GLMM. Perception of MoA was worse at word-onset position than at word-offset position (β5) for unshifted stimuli. Perception of PoA was worse than perception of MoA in the default configuration of 0 mm and word-offset position, but this difference was not statistically detectable (β9; $p = .194$). PoA perception in the shifted conditions was notably worse than performance in the default configuration (β16), but each of the successive differences between the default and the other changes (e.g., changing MoA to PoA, changing 0 to 6 mm, changing word-offset to word-onset, changing the PoA:shift interactions, etc.) tended to not reach statistical detection on their own. In other words, despite the large leap between the default performance of 99.7% and 52.5% 6 mm word-onset performance for PoA, that leap appears to have been the composite of many smaller more modest steps that collectively contribute to the score, rather than any particular large effect.

For phonemes in word-offset position, there was a detrimental effect of spectral shifting on MoA in the 4 mm condition (β3; β = –2.23). However, for the voicing feature, there was an interaction with this effect that was equivalent but in the opposite direction (β19; β = +2.1), suggesting that the voicing feature in word-offset position was not affected by spectral shifting. However, the robustness of consonant voicing against spectral shifting was restricted to the word-final position, as indicated by the negative interactions between voicing and word position (table lines 22, 23, and 24), which counteracted the positive coefficient for voicing in word-final position (table lines 18, 19, and 20). The robustness of consonant voicing perception specifically in word-offset position can be explained by the acoustic cues for voicing, such as preceding vowel duration (House 1961), which is a cue that would be robust to spectral shifting as it does not depend on any particular frequency content (Winn et al. 2012). Other specific patterns can be derived from Table 1 in the same way that we have derived the examples earlier.

Phoneme Categorization •

Group Data Overview • Figure 6A shows the group average psychometric functions for the /ʃ/-/s/ continuum across vocoder conditions. The high upper asymptotes of responses across conditions suggest listeners on average maintained consistent perception of the /s/ category, as expected. The lower asymptote of the function was maintained at or near the floor in the 0 and 2 mm conditions, but the elevated lower asymptote in the 4 and 6 mm shift conditions indicates a response bias toward /s/ even for the most /ʃ/-like sound. These results are consistent with the hypothesis, given the acoustics of /ʃ/ should sound more like /s/ when shifted upward.

Figure 6B shows the group average psychometric functions for the /b/-/d/ continuum across vocoder conditions. An expected well-defined steeply sloping function was observed in the normal (unprocessed) condition; slopes became shallower as degree of spectral shift increased, until flattening out at the 6 mm condition. This change in slope is similar to findings from earlier literature of

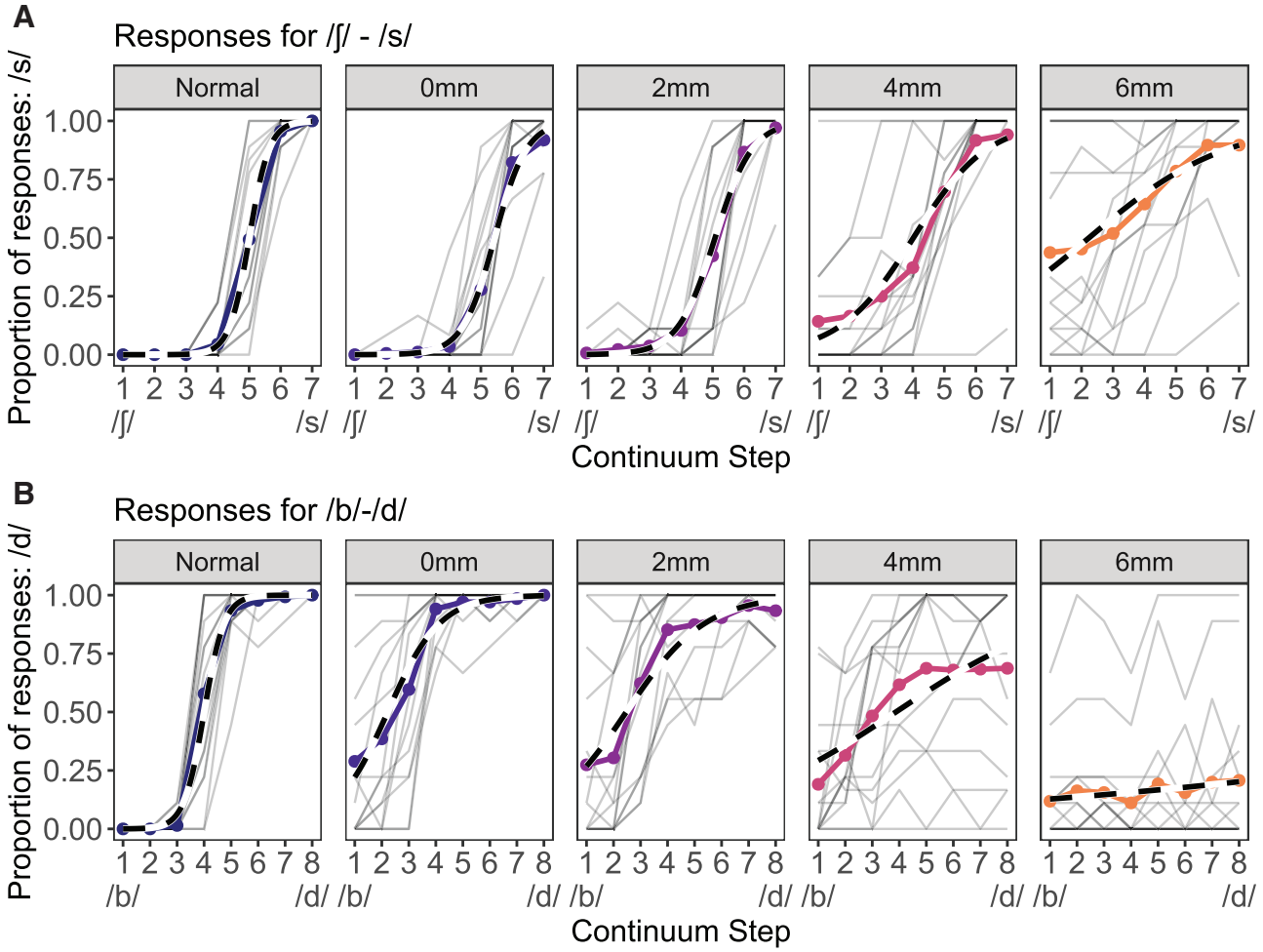

Fig. 6. Psychometric functions of phonetic categorization across both stimulus continua, separated by listening condition. The group averages are represented by the thicker lines, with individual responses in light gray. Panels in row (A) show the proportion of /s/ responses across the /ʃɑ/-/sɑ/ continuum. Panels in row (B) show proportion of /d/ responses across the /bɑ/-/dɑ/ continuum, with same line and color aesthetics as row A. Results estimated from the GLMM are represented by the dashed line. GLMM indicates generalized linear (logistic) mixed-effects model.

reduced slope resulting from poor spectral resolution overall (Winn & Litovsky 2015; Winn et al. 2016). However, the main hypothesis was not confirmed, as there was no systematic bias toward hearing /d/ with increased spectral shifting. In the 4 and 6 mm conditions, there was only a weak semblance of categorization at all, suggesting that the problem of spectral shifting was swamped by the general lack of perceptual resolution of these stimuli. Potential reasons for this result are considered in the discussion.

Generalized Linear Mixed-Effects Models for Phoneme Categorization •

/ʃa/-/sa/ • Group GLMM results for the /ʃ/-/s/ contrast are reported in Table 2. Compared with the normal listening condition (which was the default condition in the model, against which all other conditions were compared), the intercept was significantly higher for the 4 mm (β5; $z = 4.6$; $p < .001$) and 6 mm (β6; $z = 6.96$; $p < .001$) conditions, indicating a greater bias toward /s/ in each of these conditions compared with the unshifted condition, and confirming the main hypothesis. There was a significant main effect of continuum step (slope) in the normal condition (β2; $z = 8.23$; $p < .001$). Significant interactions of slope by condition were observed for the 4 mm (β9; $z = –3.92$; $p < .001$) and 6 mm (β10; $z = –6.34$; $p < .001$) conditions, where the slope was shallower.

/ba/-/da/ • GLMM results for the /b/-/d/ contrast are reported in Table 3. Consistent with previous studies of spectral distortion on categorization of these phonemes, the steepest slope of the psychometric function was observed in the normal unprocessed condition (β2; $z = 6.77$; $p < .001$). Slope Interactions were observed for the vocoder conditions, with systematically shallower slopes for the 0 mm (β7; $z = –4.82$; $p < .001$), 2 mm (β8; $z = –4.49$; $p < .001$), 4 mm (β9; $z = –5.71$; $p < .001$), and 6 mm (β10; $z = –6.93$; $p < .001$) conditions.

The main hypothesis for the /b/-/d/ stimuli was that there would be increased likelihood of perceiving /d/ when the spectrum was shifted upward. The results did not confirm the hypothesis. For most of the conditions, there was no effect on the intercept, indicating no systematic bias toward /d/. There was a negative effect on the intercept for the 6 mm shift (β6; $z = –3.91$; $p < .001$) condition. At a glance, this appears to indicate a bias toward /b/, which is the opposite of what we expected. However, this statistical effect is not as easily interpretable as the corresponding effect for the /ʃ/-/s/ stimuli. For the /b/-/d/ stimuli, the labeling functions did not show a systematic shift so much as they completely broke down.

A second GLMM was used to model only the subset of data where participant responses were matched to the continuum

TABLE 2. GLMM binomial model results for the /ʃ/-/s/ contrast

Term	Fixed Effects	Estimate	SE	*z* Statistic	*p*
	Default condition (normal speech)				
β1	Intercept	–8.93	1.21	–7.40	< .001
β2	Continuum step (slope)	4.42	0.54	8.23	< .001
	Intercept interactions				
β3	Intercept: 0 mm shift	–0.60	1.91	–0.31	.75
β4	Intercept: 2 mm shift	1.69	1.75	0.97	.33
β5	Intercept: 4 mm shift	6.39	1.38	4.63	< .001
β6	Intercept: 6 mm shift	9.79	1.41	6.96	< .001
	Slope interactions				
β7	Continuum step × 0 mm shift	–0.49	0.85	–0.58	.56
β8	Continuum step × 2 mm shift	–0.89	0.83	–1.07	.28
β9	Continuum step × 4 mm shift	–2.48	0.63	–3.92	< .001
β10	Continuum step × 6 mm shift	–3.68	0.58	–6.34	< .001

GLMM, generalized linear (logistic) mixed-effects model.

from which the stimulus was drawn (e.g., excluding response of /sɑ/ for a stimulus from the /b/-/d/ continuum). We call such a mistake a "cross-continuum error." On average, listeners' tendency to respond within the correct continuum for the /ʃɑ/ -/sɑ/ contrast ranged between 98 and 100% for all conditions, while /bɑ/-/dɑ/ ranged from 100% in the 2 mm, 0 mm, and normal conditions down to 80% in the 4 mm condition, and down to only 52% in the 6 mm condition (consistent with our earlier point that /b/-/d/ perception in the 6 mm condition is less interpretable because basic recognition of the speech sounds was so poor). Despite the occurrence of cross-continua errors in the most difficult condition, there were no changes in the statistical differences identified in model outputs of the GLMM results; the same model terms and levels reached statistical detection, with the same direction of each effect. In addition, a follow-up analysis with data subset as a model term showed no interactions of any of the model terms with data subset. All figures and model results in the current analysis are presented here using the first variation of the GLMM that included all responses.

Four-Parameter Statistical Model • Although the perceptual responses were binomial in nature, the binomial GLMM did not explicitly model the true lower or upper asymptotes within the data collected, due to the mathematical constraint that the data would extrapolate to 0 and 1. As can be seen from the data displayed in Figure 6, the range from lower to upper asymptote varies across the conditions. We are using the range between these asymptotes as an index of a listener's tendency to show balanced perception of two phonetic categories, as opposed to the dominance of one category. Balanced and successfully recalibrated perception would manifest as lower and upper asymptotes fully separated at 0 and 1, respectively, which indicate that there are acoustic levels within the continuum that are reliably heard as each of the phonemes. Conversely, if responses cluster near the floor or ceiling of the graph, that is a sign that only one of the phonemes is being identified reliably. To directly model changes in response range, a four-parameter model was used, in which the standard parameters of slope and crossover boundary were supplemented with parameters that estimated lower and upper asymptote.

/ʃɑ/-/sɑ/ • Results of the four-parameter sigmoidal function for the /ʃɑ/-/sɑ/ contrast are listed in Table 4. The most important aspect of this table is the "Range" parameter. The Estimate of –0.51 in the 6 mm condition means that the average range was 0.49, which is 1.0 (the estimate for the normal-speech condition) minus 0.51. Compared with the normal condition, the range of responses was smaller in the 4 mm (t = –3.58; p < .001) and 6 mm (t = –8.24; p < .001) condition but was not statistically different in the 0 mm (t = –1.71; p = .09) or 2 mm (t = –0.59; p = 0.56) conditions. These results confirm the main hypothesis that there would be successively greater biases

TABLE 3. GLMM binomial model results for the /b/-/d/ contrast

Term	Fixed Effects	Estimate	SE	*z* Statistic	*p*
	Default condition (normal speech)				
β1	Intercept	2.70	0.82	3.28	.001
β2	Continuum step (slope)	4.73	0.70	6.77	< .001
	Intercept interactions				
β3	Intercept: 0 mm shift	0.99	0.71	1.39	.17
β4	Intercept: 2 mm shift	0.53	1.06	0.50	.62
β5	Intercept: 4 mm shift	–1.73	1.08	–1.60	.11
β6	Intercept: 6 mm shift	–6.54	1.67	–3.91	< .001
	Slope interactions				
β7	Continuum step × 0 mm shift	–3.04	0.63	–4.82	< .001
β8	Continuum step × 2 mm shift	–3.16	0.71	–4.49	< .001
β9	Continuum step × 4 mm shift	–3.95	0.69	–5.71	< .001
β10	Continuum step × 6 mm shift	–5.00	0.72	–6.93	< .001

GLMM, generalized linear (logistic) mixed-effects model.

TABLE 4. Four-parameter model output for the /ʃ/-/s/ contrast

Term	Parameter	Condition	Estimate	SE	*t* Statistic	*p*
β1	Range	Normal	1.00	0.03	29.34	< .001
β2		0 mm	–0.08	0.05	–1.71	.09
β3		2 mm	–0.03	0.05	–0.59	.56
β4		4 mm	–0.19	0.05	–3.58	< .001
β5		6 mm	–0.51	0.06	–8.24	< .001
β6	Floor	Normal	0.00	0.02	–0.02	.98
β7		0 mm	0.01	0.02	0.42	.68
β8		2 mm	0.02	0.02	0.90	.37
β9		4 mm	0.15	0.03	5.79	< .001
β10		6 mm	0.43	0.03	12.97	< .001
β11	Slope	Normal	3.01	0.59	5.14	< .001
β12		0 mm	–0.10	0.73	–0.13	.90
β13		2 mm	–0.85	0.66	–1.28	.20
β14		4 mm	–1.27	0.64	–1.98	< .05
β15		6 mm	–1.69	0.68	–2.47	.01
β16	Midpoint offset	Normal	–5.02	0.04	–113.38	< .001
β17		0 mm	–0.27	0.08	–3.43	< .001
β18		2 mm	–0.13	0.08	–1.56	.12
β19		4 mm	0.39	0.10	3.74	< .001
β20		6 mm	0.83	0.21	4.05	< .001

TABLE 5. Four-parameter model output for the /b/-/d/ contrast

Term	Parameter	Condition	Estimate	SE	*t* Value	*p*
β1	Range	Normal	0.99	0.03	34.17	< .001
β2		0 mm	–0.30	0.05	–6.54	< .001
β3		2 mm	–0.33	0.05	–6.56	< .001
β4		4 mm	–0.41	0.10	–4.26	< .001
β5		6 mm	–0.45	0.01	–0.02	.99
β6	Floor	Normal	–0.01	0.02	–0.36	.72
β7		0 mm	0.31	0.04	8.37	< .001
β8		2 mm	0.26	0.04	6.43	< .001
β9		4 mm	0.14	0.09	1.64	.10
β10		6 mm	0.13	0.19	0.67	.50
β11	Slope	Normal	3.50	0.87	4.02	< .001
β12		0 mm	–1.10	1.01	–1.09	.28
β13		2 mm	–1.21	1.01	–1.19	.23
β14		4 mm	–2.17	0.95	–2.29	.02
β15		6 mm	–3.20	3.04	–1.05	.29
β16	Midpoint offset	Normal	–3.91	0.05	–86.10	< .001
β17		0 mm	0.84	0.10	8.76	< .001
β18		2 mm	0.98	0.11	8.98	< .001
β19		4 mm	1.37	0.30	4.56	< .001
β20		6 mm	–9.53	0.67	–0.04	.97

toward the higher-frequency /s/ sound with increasing upward spectral shift. The other effects listed in Table 3 were mainly consistent with the binomial GLMM.

/ba/-/da/ • Results for the /bɑ/-/dɑ/ contrast are listed in Table 5. Compared with the normal condition, the range of responses was significantly smaller for each of the 0 mm (t = –6.54; p < .001), 2 mm (t = –6.56; p < .001), and 4 mm (t = –4.26; p < .001) conditions but without any systematic effect across conditions. In other words, all vocoder conditions were different from the unprocessed condition, but the 0, 2, and 4 mm conditions were not meaningfully different from each other. The results from the 6 mm condition are especially difficult to interpret; upon inspection of the individual models, it was found that some listeners produced identification functions that were so flat that they were estimated to not reach 50% until far past the limits of the continuum, and indeed, past any realistic acoustic constraints. For this reason, the statistical results for the 6 mm condition are not described here.

Individual Differences in Recalibrating to Spectral Shifts • The core value of the current analysis is appreciating how perception of phonemes—and the recalibration of acoustic-to-phonetic mapping—is subject to individual differences. Figure 7 shows examples of starkly different patterns of individual results for the /ʃ/-/s/ contrast from four listeners. Listeners A and B showed reliable identification of both /ʃ/ and /s/ in all conditions, regardless of spectral shift. It thus appears that they successfully adjusted (recalibrated) their acoustic-phonetic boundaries for these phonemes commensurate with the spectral shifting. Conversely, listeners C and D showed a significant bias toward /s/ in the 4 and 6 mm condition, indicated by the increased floor of the lower asymptote of the psychometric function. In other words, for listeners C and D categorization of the /ʃ/ tokens were shifted into the acoustic region normally corresponding to /s/, and their perceptual boundary did not shift accordingly. These listeners were not successful recalibrators.

There was relatively more variability observed for the /b/-/d/ contrast than for the /ʃ/-/s/ contrast. Some listeners showed a categorization bias toward /d/ in the 0, 2, and 4 mm conditions. However, /b/-/d/ perception in the 6 mm condition was extremely difficult for all listeners, with some biased either completely toward /d/, or surprisingly, completely toward /b/. These patterns of responses are likely indicative of categorization breaking down and listeners simply guessing because the stimuli were so difficult to distinguish.

Two Distinct Patterns of Recalibration • In light of the distinctly different patterns of perceptual recalibration described earlier and illustrated in Figure 7, follow-up analyses were conducted to quantify the extent of recalibration to acoustic-phonetic mapping in response to spectral shifting. In general, listeners clustered into one of two groups; some could maintain perceptual separation of both sound categories even when spectra were shifted, while others were heavily biased to generally perceive only one sound category (e.g., Listeners C and D in Fig. 7, for whom upward-shifted fricatives would almost always sound like /s/). The range between floor and ceiling for the psychometric function for each listener was used as a proxy for a listener's ability to recalibrate to the spectral shift. Results are illustrated for the /ʃ/-/s/ and /b/-/d/ contrasts in Figure 8. A key finding of the individual results is that about half of the listeners had a full or near-full response range (indicating that they could recalibrate to a spectral shift and maintain two separate phonetic categories), while the other half did not.

The individual analysis of response ranges in the heavily shifted conditions revealed that the average group response was not representative of any individual listener. Instead, for both the stop and fricative contrasts, listeners were split into a bimodal distribution of those who could and who could not recalibrate to varying amounts of spectral shifting. This result is illustrated in Figure 8 in the 6 mm condition for /ʃ/-/s/, and in the 4 mm condition for /b/-/d/, where the individual listener data (small black dots) are clustered together at either the top or

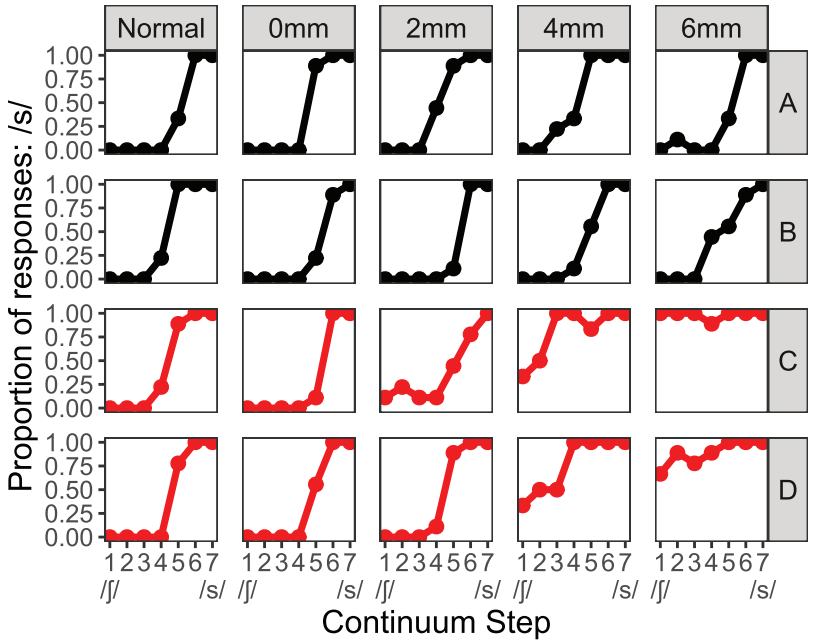

Fig. 7. Individual psychometric functions of four unique listeners for the /ʃɑ/-/sɑ/ continuum. Columns correspond to the five different listening conditions that were tested (normal, 0, 2, 4, and 6 mm of spectral shift), with individual listeners in each row. Listeners A and B had steeply sloping response functions that rose from 0 to 1 in each condition, indicating they maintained two phonetic categories despite spectral shifting. Listeners C and D were heavily biased toward /s/ in the conditions with greater spectral shift, with shallow response slopes and a rising floor of the function.

bottom of the plot, corresponding to successful or unsuccessful recalibration, respectively, despite the group average indicating something in between these extremes.

DISCUSSION

The impact of spectral shifting on perception of speech was tested to better understand its impact on overall accuracy and specific acoustic-to-phonetic mapping. The pattern of results for word recognition and phoneme categorization indicated an expected decrease in perceptual accuracy when the spectrum is shifted. However, the most significant finding of the present study was the effect of frequency shifting on speech recognition was not uniform across listeners. Recalibration to spectral shifting was operationally defined as changing acoustic-phonetic category mapping commensurate with the amount of frequency shifting, rather than demonstrating bias toward only hearing sounds as having high-frequency components. There were listeners who could recalibrate to the spectral shift or listeners who failed to recalibrate altogether, showing inability to adjust their acoustic-phonetic mapping (Figs. 7 and 8). There did not appear to be any evidence of listeners with results intermediate to these two patterns. The ability to recalibrate to a spectral shift was not as clear when analyzing performance of CNC scores alone or even with a phonetic feature analysis. Instead, this ability only emerged with a more detailed analysis of phoneme categorization that was directly driven by the shifted frequency cues.

Even though word recognition is a standard and understandable performance metric, it is dependent on a multitude of factors (spectro-temporal cues, lexical knowledge, etc.), any of which could obscure the specific effect of adaptation to frequency shifting. In the present study, we supplemented word-recognition testing with a specific phonetic categorization task where the outcome measure was directly related to whether the listener could successfully remap frequencies to phonemes. Phonemes /ʃ/ and /s/ are examples of speech sounds where shifting frequencies should have a specific impact on categorization, as shifting /ʃ/ upward in frequency should render it more /s/-like. However, these phonemes are not necessarily contrasted by absolute frequency peaks, as the spectra of these sounds can also be described as differing by relative energy across multiple peaks. Specifically, the /s/ phoneme might therefore be defined not by its absolute peak but by virtue of its upper peak being stronger than its lower peak (and vice versa for /ʃ/). These spectral properties are captured as asymmetry, or alternatively as relative changes in energy in the same frequency regions across the fricative-vowel boundary, which has shown to be a powerful explanation of fricative perception in CI users (Hedrick & Carney 1997). Therefore, it is possible that the inability to recalibrate to spectral shifting could represent the difficulty of tracking the spectral asymmetry when it is shifted, rather than the difficulty in moving a perceptual boundary.

The main hypothesis of this study was validated for the /ʃ/-/s/ contrast but not for the /b/-/d/ contrast. Upward frequency shifting promoted greater bias toward /s/ (Fig. 6A) but did not promote bias toward /d/ (Fig. 6B). One potential reason is the difference in acoustic cues that are used when distinguishing these phoneme pairs. The /bɑ/-/dɑ/ contrast relies on distinction of formant peaks within an information-rich and relatively narrow spectral bandwidth (less than one octave), as opposed to the fricatives, which occupy over two octaves of frequency space, and whose peak frequencies are more diffuse. Therefore, the very presence of spectral degradation should be more detrimental for /b/-/d/ than for /ʃ/-/s/, even before any frequencies are shifted. The formant transition cues necessary to perceive /b/-/d/ only last for approximately 60 to 80 ms before converging into the vowel,

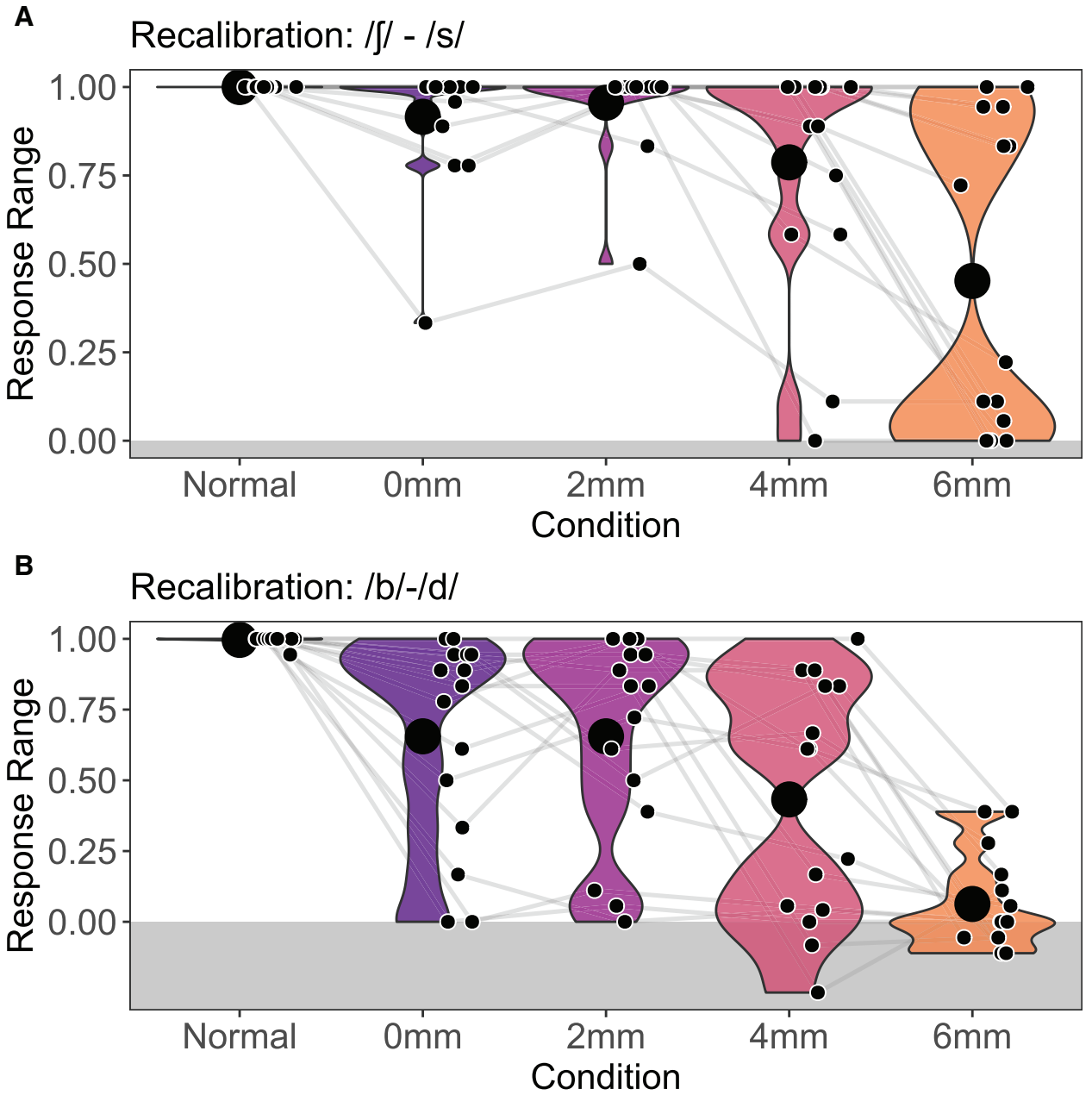

Fig. 8. Response ranges of psychometric functions in each listening condition for each continuum. Labeling function response range (separation of lower and upper asymptotes) in each listening condition for (A) the /ʃ/-/s/ and (B) /b/-/d/ contrast. The large black dot represents the group mean, with individual data points represented by the small black dots. Lines connect data for individuals across conditions. Width of the violin shape underneath the dots reflects the density distribution of the individual data. The gray shaded region represents response range values that were negative, that is, the asymptote for the lower end of the continuum was paradoxically higher than the asymptote for the upper end of the continuum.

compared with the relatively longer duration of fricative noise (about 180 ms). Past studies corroborate this, with categorization of /ʃ/-/s/ by CI listeners being rather similar to that of listeners with NH (Lane et al. 2007; Winn et al. 2013), while /b/-/d/ perception is relatively poorer (Winn & Litovsky 2015). The effects of shifting for /b/-/d/ appear to result mainly from signal degradation (i.e., vocoding) and not a spectral shift per se. With these considerations in mind, we contend that categorization of /ʃ/-/s/ fricative sounds would be a useful probe for perceptual adaptation to frequency-shifted speech, because they do not force the listener to reckon with other detrimental signal degradations that would severely impact perception of the contrast.

Phonetic Feature Perception

Some phonetic features are more susceptible to misperception when the spectrum is shifted. Voicing features were robust to the spectral shift, particularly for consonants in word-final position. This almost certainly is attributable to the fact that voicing in that position is cued reliably by vowel duration (House 1961), which is a property essentially unchanged by the spectral shift, and in fact, robust to background noise (Revoile et al. 1986), sensorineural hearing loss (Owens 1978) and cochlear implantation (Winn et al. 2012). Manner and PoA were more difficult to perceive (Fig. 5). PoA has consistently been shown to be a difficult feature for CI listeners to perceive (Munson et al. 2003;

Winn & Litovsky 2015), and given the results of the present study, a frequency mismatch plays a significant role in accurately perceiving this cue. Perceiving PoA requires accurate perception of spectral properties, and thus this feature is less robust to spectral degradation and spectral shifting, regardless of word position. The categorization task and the word recognition task both showed that misperception of frequency-shifted phonemes can reflect systematic changes in perceived PoA.

Speculating on the Source of Individual Differences

The bimodal distribution of performance for recalibration invites the question of what makes an individual capable or incapable of adjusting their individual acoustic-phonetic mapping. Among this sample of NH listeners, it is customary to assume that peripheral sensitivity was sufficiently similar across the sample of listeners, implying that the capacity to recalibrate hinges on some higher-level cognitive ability, such as being able to quickly adapt to rule switching. Rosemann et al. (2017) found several cognitive factors to account for noise-vocoded speech perception in both young and older adults including vocabulary size, working memory, and task switching abilities. However, Erb et al. (2012) found that adaptation to four-channel noise-vocoded speech was explained better by performance on nonspeech tasks (amplitude modulation discrimination) more than working memory. In a follow-up study, functional magnetic resonant imaging results showed when listening to four-channel noise-vocoded speech, listeners rely on a higher-level executive network of cortical activation, with rapid adaptation to degraded speech being only partly regulated by top-down networks (Erb et al. 2013). It is worth noting the aforementioned studies did not use spectrally shifted speech, but it is feasible to speculate that the challenge of adapting to a degraded input might invoke common compensatory mechanisms. There is precedent for that line of reasoning in the scientific literature on speech motor control, where control of ballistic arm movements is compared with the control of motor movements required to produce speech (Max et al. 2003). In addition, it is possible listeners who were successful recalibrators were able to use a different cue when the primary spectral cue needed to perceive changes in PoA was no longer reliable in vocoded and spectrally shifted conditions. It is likely that listeners relied on the formant cue in the unprocessed condition; however, it remains possible for listeners to have variation in their perceptual strategies, particularly in the more challenging conditions.

It would be reasonable to suspect that the ability to accommodate spectral shifting is related to one's ability to accommodate variation in talker vocal tract size, as the impact of changes in vocal tract length are expressed in proportional shifting of formant frequencies. In the present study, we defined the degree of spectral shifts in terms of cochlear space; converting these shifts to proportions results in roughly 40%, 92%, and 161% for the 2, 4, and 6 mm conditions, respectively. These frequency shifts far exceed the frequency changes resulting from differences in vocal tract size across talkers of typical size. The difference in formant frequencies between vowels produced by women and men are roughly on the order of 15 to 20% (Fant 1966; Hillenbrand et al. 1995; Goldstein Reference Note 1). Therefore, the adaptation to different talkers' vocal tracts in everyday life is not comparable to adapting to spectral shifts of several millimeters in cochlear space, as would be needed when listening through a CI. One additional detail to consider is that the source voice used for these stimuli was an adult male, meaning that upward frequency shifts passed through a frequency region that would be typical for an adult female or a child. If the stimuli were spoken by a woman or a child, it is possible that even less shifting could be accommodated, as the vocal tract size would be already near the upper end of what would typically be encountered.

Training and Adaptation

Listeners in the present study were acutely exposed to spectral shifts, and it is not known whether the results found here would persist over the course of prolonged experience with shifted speech. Rosen et al. (1999) found that in a group of NH listeners, perception of consonants, vowels, and words with a 6.5 mm shift in cochlear space improved significantly after nine 20-minute training sessions. However, it was unclear if the results were from mere exposure to spectrally shifted speech or if listeners would gain more benefit from explicit training. Fu et al. (2005) found that over 5 consecutive days, performance improved with training rather than with mere exposure, and was improved mainly for the utterance type that was trained (e.g., phonemes, not sentences). Exposure to spectral shifting also improves perception of sentence-length stimuli (Fu & Galvin 2003; Nogaki et al. 2007) and has shown benefits that are specific to spectral shifting rather than spectral degradation in general (Faulkner et al. 2012).

The full-time course of adaptation to spectrally shifted speech is not completely known. More successful adaptation to spectrally shifted speech is observed when the shift is introduced gradually rather than suddenly (Svirsky et al. 2015b). Svirsky et al (2004) tested CI listeners, who must concurrently adapt to both frequency-degraded and frequency-shifted signals, finding that it took a period of time ranging from 0 days to 2 years before performance reached plateau. It is interesting that CI listeners tested by Fu and Shannon (1999) showed the same sensitivity to spectral shifting as NH listeners, despite the two listener groups having dramatically different amounts of experience with degraded/shifted speech (acute exposure for NH versus everyday experience with CI). The collection of results described earlier suggests that more listeners in the present study could have recalibrated their acoustic-phonetic mapping if training were given, but the exact timeline of recalibration is still largely unknown.

Clinical Implications

Individual variability is a known feature of speech perception performance among CI users, with many contributing factors (Lazard et al. 2012; Blamey et al. 2013; Holden et al. 2013). Some of the variability is likely due to variation in insertion depth in individual CIs. The present study further suggests that in addition to the variability of insertion depth, there is an additional layer of variability attributable to the listener's capacity to adjust to the resulting spectral shift. That is, two CI recipients could have the same degree of insertion depth and yet have very different abilities to accommodate the frequency-place mismatch. This has implications for newly implanted recipients, who could be managed differently based on their capacity to adjust to spectrally shifted speech. Specifically, those who can more easily handle a spectral shift can likely adapt to the default frequency-channel allocation of the device, which sacrifices tonotopy in favor of wider coverage of the frequency spectrum. Conversely, there could be some implant

recipients for whom preservation of accurate tonotopy would be more beneficial than wide-spectrum coverage; for such listeners, the disadvantage of excluding low-frequency energy would be counteracted by a relatively larger advantage of preserving frequency-place match.

In line with the clinical goals stated earlier, Fitzgerald et al. (2013) tested whether adjustments to frequency-channel allocation could be guided by the individual listener. NH listeners performed real-time adjustments to the frequency allocation for spectrally shifted carrier channels, until speech was "most intelligible" to them. They found significantly better CNC word recognition for listeners using the self-selected frequency tables that were tonotopically aligned with the noise bands compared with the other standard mapping tables covering a wider frequency range but with some frequency mismatch. In a follow-up study, Fitzgerald et al. (2017) showed that in cases where both ears have unequal tonotopic mismatch (as is the case for many bilateral CI recipients), listeners prefer a frequency-channel allocation that minimizes the frequency mismatch between ears.

While there has been some work examining the effects of frequency allocation table adjustments in CI recipients (Svirsky et al. 2015a), further research is needed to fully understand the clinical feasibility of the self-selection approach for frequency allocation. Given the results of the present study—even acknowledging that it used NH listeners—we contend that the specific benefit or detriment of tonotopic mismatch should be tested with stimuli where performance specifically depends on the ability to recalibrate acoustic-phonetic mapping.

Limitations

The present study has some methodological limitations that are worth noting. First, vocoded speech is only a crude approximation of some aspects of CI signal processing and cannot replicate the impact of experience with the device and/or atrophy of the auditory system. Most importantly, the tonotopic mismatch used here was simplified relative to what is actually observed in a real CI. It has become common to express insertion as an angle (number of degrees of cochlear turns) rather than a mm distance. A comprehensive study by Landsberger et al. (2015) suggests that the mean insertion angle varies across devices with angles of 391 (694 Hz), 375 (740 Hz), 561 (323 Hz), and 486 (467 Hz) for the AB Hi-Focus 1J, Cochlear Contour Advance, Med-El standard, and Med-El Flex28, respectively. The amount of frequency mismatch also would vary across devices for a second reason, which is that each processor has a different frequency-channel allocation and also different interelectrode spacing (James et al. 2019). To complicate matters further, the amount of frequency mismatch is not constant across the entire array, with greater mismatch at the apex (where the angle measurements are estimated) than at the base.

The default center frequencies for the most apical channel for Cochlear is 242 Hz, 322 Hz for Advanced Bionics, and 149 Hz for Med-El's fine structure processing strategy. These frequencies correspond to 6.16, 7.52, and 4.18, respectively, along the basilar membrane. However, the mm scale derived with the Greenwood function is not suitable for spiral ganglion stimulation. Using estimates from Stakhovskaya et al. (2007) and Landsberger et al. (2015), the aforementioned default most apical frequencies correspond to angular insertions of 610, 562, and 655 degrees, respectively. Incorporating their spiral ganglion modeling and the standard frequency allocation for the most apical electrode of each device, we estimate the following amounts of frequency mismatches of 372 Hz (~0.87 octaves) for the Advanced Bionics Hi-Focus 1J, 498 Hz (~0.49 octaves) for the Cochlear Contour Advance, 174 Hz (~0.86 octaves) for the Med-El standard array, and 318 Hz (~0.47 octaves) for the Med-El Flex28.

In the phoneme categorization component of the present study, listeners were constrained to respond using one of only six choices. It is possible the listener would have preferred to respond outside of this response range. For example, maybe a listener heard /nɑ/ when listening to a /bɑ-dɑ/ token, and only selected /bɑ/ or /dɑ/ because this was closer to what they heard compared with the other four choices. In a meta-analysis, Rødvik et al. (2018) saw that CI listeners confused /b/ with /n/ more than /d/, and it is possible listeners in the present study had a similar confusion but were unable to respond with what they actually heard. Being able to measure what the listener actually thinks they hear is more desirable than showing whether they picked the best match from available choices, although closed-set tasks are simpler to analyze and can help focus the search for perception of specific cues.

The participants in this study were confirmed to have NH thresholds up to 8 kHz, yet in several conditions, there were frequency carrier channels higher than 8 kHz (Fig. 1). It is therefore possible that extended high-frequency hearing was a factor in the results. However, there are two reasons why we believe this explanation is not powerful enough to explain the current results. First, the sounds presented here were far above threshold level. Second, and more importantly, the effect described here is not about detection of a high-frequency signal but rather about distinguishing two different spectral patterns. The most important part of the study was the perception of the lower-frequency phoneme /ʃ/, which fell into an audible frequency range even when shifted upward. Even when bandwidth is limited, perceptual distinction of /ʃ/ and /s/ is still possible (Miller & Nicely 1955; Alexander & Rallapalli 2017).

A final point that remains unaddressed by this study and nearly all other studies is how the impact of spectral shifting interacts with the challenge of noise masking. Listening to speech in noise is often the chief concern of individuals with hearing impairment, including those who use CIs. Future studies might find that the results obtained in the present study are either independent of the difficulty of noise masking or perhaps exacerbate the problem.

CONCLUSIONS

When the spectrum of speech is shifted higher in frequency, listeners suffer poorer perception of speech overall, with particular difficulty on consonant PoA. In some cases, there is a bias toward hearing phonemes with higher-frequency spectral peaks, as if the acoustic-phonetic mapping is not completely proportionate to the upward shift of frequency. Half of the listeners were able to successfully recalibrate to the spectrally shifted stimuli by adjusting their acoustic-phonetic mapping commensurate with the shifting of the input, while the other half of the listeners did not. The larger implication of this finding is that the difficulties in adjusting to shallow insertion depth—which is a very common feature of CIs—is not predictable based solely on

the amount of tonotopic mismatch. Instead, it is also subject to substantial individual differences in the ability to accommodate to a shifted input. Group-averaged data in studies of spectrally shifted speech might imply that we should expect the typical listener to show partial adjustment to spectral shift, but the current individual-level analyses suggest instead that there are individuals who fully adjust and individuals who essentially fail to adjust, with few in between.

The clustering of the listener groups was more pronounced in the phoneme categorization test compared with the word-recognition task. This makes sense because typical CNC words are contrasted by more than just the spectral domain. That is, there are many ways to perceive or misperceive the stimulus, while the categorization task specifically depended on acoustic-phonetic remapping in the frequency domain. Together with previous methods of probing for individuals' preference for frequency-channel allocation, the current approach might be helpful in identifying CI recipients who have specific deficits in accommodating frequency-shifted input that are unexplained by routine auditory measures. That information can be used to tailor CI map parameters (frequency-electrode allocation) to suit an individual's need for tonotopic match.

ACKNOWLEDGMENTS

The authors are grateful to Ashley N. Moore and Moira McShane for their assistance with data collection.

This work was supported by National Institutes of Health National Institute on Deafness and Other Communication (NIDCD) R03 DC 014309 (M.B.W.) and NIDCD R01 DC017114 (M.B.W.).

Portions of this article were presented as a poster at the 2018 fall meeting of the Acoustical Society of America (November 5–9, 2018, Victoria, British Columbia, Canada) and the Conference on Implantable Auditory Prostheses 2019 (July 14–19, 2019, Lake Tahoe, California, USA).

The authors have no conflicts of interest to disclose.

Address for correspondence: Michael L. Smith, Department of Speech-Language-Hearing Sciences, University of Minnesota, 164 Pillsbury Dr SE, Minneapolis, MN 55455, USA. E-mail: smit8854@umn.edu

Received March 20, 2020; accepted January 25, 2021

REFERENCES

Alexander, J. M., & Rallapalli, V. (2017). Acoustic and perceptual effects of amplitude and frequency compression on high-frequency speech. *J Acoust Soc Am*, *142*, 908.

American National Standards Institute Accredited Standards Committee S3, Bioacoustics. (2004). *American National Standard Methods for Manual Pure-Tone Threshold Audiometry*. Standards Secretariat, Acoustical Society of America.

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J Stat Softw*, *67*, 1–48.

Bierer, J. A., Faulkner, K. F., Tremblay, K. L. (2011). Identifying cochlear implant channels with poor electrode-neuron interfaces: Electrically evoked auditory brain stem responses measured with the partial tripolar configuration. *Ear Hear*, *32*, 436–444.

Bierer, J. A., & Litvak, L. (2016). Reducing channel interaction through cochlear implant programming may improve speech perception: Current focusing and channel deactivation. *Trends Hear*, *20*, 2331216516653389.

Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., Dillier, N., Dowell, R., Fraysse, B., Gallégo, S., Govaerts, P. J., Green, K., Huber, A. M., Kleine-Punte, A., Maat, B., Marx, M., Mawman, D., Mosnier, I., O'Connor, A. F., O'Leary, S., et al. (2013). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients. *Audiol Neurootol*, *18*, 36–47.

Boersma, P., & Weenink, D. (2013). Praat: Doing Phonetics by Computer [Computer program]. Version 5.3.56. http://www.fon.hum.uva.nl/praat/.

DiNino, M., Wright, R. A., Winn, M. B., Bierer, J. A. (2016). Vowel and consonant confusions from spectrally manipulated stimuli designed to simulate poor cochlear implant electrode-neuron interfaces. *J Acoust Soc Am*, *140*, 4404.

Dorman, M. F., Loizou, P. C., Rainey, D. (1997). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *J Acoust Soc Am*, *102*(5 Pt 1), 2993–2996.

Dorman, M. F., Loizou, P. C., Fitzke, J., Tu, Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels. *J Acoust Soc Am*, *104*, 3583–3585.

Erb, J., Henry, M. J., Eisner, F., Obleser, J. (2012). Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. *Neuropsychologia*, *50*, 2154–2164.

Erb, J., Henry, M. J., Eisner, F., Obleser, J. (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *J Neurosci*, *33*, 10688–10697.

Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report*, *1*, 22–30.

Faulkner, A., Rosen, S., Green, T. (2012). Comparing live to recorded speech in training the perception of spectrally shifted noise-vocoded speech. *J Acoust Soc Am*, *132*, EL336–EL342.

Fitzgerald, M. B., Sagi, E., Morbiwala, T. A., Tan, C. T., Svirsky, M. A. (2013). Feasibility of real-time selection of frequency tables in an acoustic simulation of a cochlear implant. *Ear Hear*, *34*, 763–772.

Fitzgerald, M. B., Prosolovich, K., Tan, C. T., Glassman, E. K., Svirsky, M. A. (2017). Self-selection of frequency tables with bilateral mismatches in an acoustic simulation of a cochlear implant. *J Am Acad Audiol*, *28*, 385–394.

Fu, Q. J., & Galvin, J. J. 3rd. (2003). The effects of short-term training for spectrally mismatched noise-band speech. *J Acoust Soc Am*, *113*, 1065–1072.

Fu, Q. J., Nogaki, G., Galvin, J. J. 3rd. (2005). Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation. *J Assoc Res Otolaryngol*, *6*, 180–189.

Fu, Q. J., & Shannon, R. V. (1999). Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *J Acoust Soc Am*, *105*, 1889–1900.

Fu, Q. J., Shannon, R. V., Galvin, J. J. 3rd. (2002). Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. *J Acoust Soc Am*, *112*, 1664–1674.

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

Harnsberger, J. D., Svirsky, M. A., Kaiser, A. R., Pisoni, D. B., Wright, R., Meyer, T. A. (2001). Perceptual "vowel spaces" of cochlear implant users: Implications for the study of auditory adaptation to spectral shift. *J Acoust Soc Am*, *109*(5 Pt 1), 2135–2145.

Hedrick, M. S., & Carney, A. E. (1997). Effect of relative amplitude and formant transitions on perception of place of articulation by adult listeners with cochlear implants. *J Speech Lang Hear Res*, *40*, 1445–1457.

Hillenbrand, J., Getty, L. A., Clark, M. J., Wheeler, K. (1995). Acoustic characteristics of American English vowels. *J Acoust Soc Am*, *97*(5 Pt 1), 3099–3111.

Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., Gotter, B. D., Vanderhoof, S. S., Mispagel, K., Heydebrand, G., Skinner, M. W. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear Hear*, *34*, 342–360.

House, A. S. (1961). On vowel duration in English. *J Acoust Soc Am*, *33*, 1174–1178.

James, C. J., Karoui, C., Laborde, M. L., Lepage, B., Molinier, C. É., Tartayre, M., Escudé, B., Deguine, O., Marx, M., Fraysse, B. (2019). Early sentence recognition in adult cochlear implant users. *Ear Hear*, *40*, 905–917.

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J. (1997). Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations. *Acta Acust United Acust*, *83*, 659–669.

Landsberger, D. M., Svrakic, M., Roland, J. T. Jr, Svirsky, M. (2015). The relationship between insertion angles, default frequency allocations, and spiral ganglion place pitch in cochlear implants. *Ear Hear*, *36*, e207–e213.

Lane, H., Denny, M., Guenther, F. H., Hanson, H. M., Marrone, N., Matthies, M. L., Perkell, J. S., Stockmann, E., Tiede, M., Vick, J., Zandipour, M.

(2007). On the structure of phoneme categories in listeners with cochlear implants. *J Speech Lang Hear Res*, *50*, 2–14.

Lazard, D. S., Vincent, C., Venail, F., Van de Heyning, P., Truy, E., Sterkers, O., Skarzynski, P. H., Skarzynski, H., Schauwers, K., O'Leary, S., Mawman, D., Maat, B., Kleine-Punte, A., Huber, A. M., Green, K., Govaerts, P. J., Fraysse, B., Dowell, R., Dillier, N., Burke, E., et al. (2012). Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: A new conceptual model over time. *PLoS One*, *7*, e48739.

Li, T., Galvin, J. J. 3rd, Fu, Q. J. (2009). Interactions between unsupervised learning and the degree of spectral mismatch on short-term perceptual adaptation to spectrally shifted speech. *Ear Hear*, *30*, 238–249.

Li, T., & Fu, Q. J. (2010). Effects of spectral shifting on speech perception in noise. *Hear Res*, *270*, 81–88.

Max, L., Caruso, A. J., Gracco, V. L. (2003). Kinematic analyses of speech, orofacial nonspeech, and finger movements in stuttering and nonstuttering adults. *J Speech Lang Hear Res*, *46*, 215–232.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *J Acoust Soc Ame*, *27*, 338–352.

Munson, B., Donaldson, G. S., Allen, S. L., Collison, E. A., 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.

Munson, B., & Nelson, P. B. (2005). Phonetic identification in quiet and in noise by listeners with cochlear implants. *J Acoust Soc Am*, *118*, 2607–2617.

Nogaki, G., Fu, Q. J., Galvin, J. J. 3rd. (2007). Effect of training rate on recognition of spectrally shifted speech. *Ear Hear*, *28*, 132–140.

Oxenham, A. J., & Kreft, H. A. (2014). Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing. *Trends Hear*, *18*, 2331216514553783.

Owens, E. (1978). Consonant errors and remediation in sensorineural hearing loss. *J Speech Hear Disord*, *43*, 331–347.

Peterson, G. E., & Lehiste, I. (1962). Revised CNC lists for auditory tests. *J Speech Hear Disord*, *27*, 62–70.

R Core Team. (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/

Revoile, S. G., Holden-Pitt, L. D., Edward, D. M., Pickett, J. M. (1986). Some rehabilitative considerations for future speech-processing hearing aids. J Rehabil Res Dev, 23, 89–94.

Rødvik, A. K., von Koss Torkildsen, J., Wie, O. B., Storaker, M. A., Silvola, J. T. (2018). Consonant and vowel identification in cochlear implant users measured by nonsense words: A systematic review and meta-analysis. *J Speech Lang Hear Res*, *61*, 1023–1050.

Rødvik, A. K., Tvete, O., Torkildsen, J. V. K., Wie, O. B., Skaug, I., Silvola, J. T. (2019). Consonant and vowel confusions in well-performing children and adolescents with cochlear implants, measured by a nonsense syllable repetition test. *Front Psychol*, *10*, 1813.

Rosemann, S., Gießing, C., Özyurt, J., Carroll, R., Puschmann, S., Thiel, C. M. (2017). The contribution of cognitive factors to individual differences in understanding noise-vocoded speech in young and older adults. *Front Hum Neurosci*, *11*, 294.

Rosen, S., Faulkner, A., Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *J Acoust Soc Am*, *106*, 3629–3636.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Lang Linguistics Compass*, *3*, 621–640.

Stakhovskaya, O., Sridhar, D., Bonham, B. H., Leake, P. A. (2007). Frequency map for the human cochlear spiral ganglion: Implications for cochlear implants. *J Assoc Res Otolaryngol*, *8*, 220–233.

Stone, M. A., Moore, B. C., Greenish, H. (2008). Discrimination of envelope statistics reveals evidence of sub-clinical hearing damage in a noise-exposed population with 'normal' hearing thresholds. *Int J Audiol*, *47*, 737–750.

Story, B. H., Vorperian, H. K., Bunton, K., Durtschi, R. B. (2018). An age-dependent vocal tract model for males and females based on anatomic measurements. *J Acoust Soc Am*, *143*, 3079.

Svirsky, M. A., Fitzgerald, M. B., Sagi, E., Glassman, E. K. (2015a). Bilateral cochlear implants with large asymmetries in electrode insertion depth: Implications for the study of auditory plasticity. *Acta Otolaryngol*, *135*, 354–363.

Svirsky, M. A., Silveira, A., Neuburger, H., Teoh, S. W., Suárez, H. (2004). Long-term auditory adaptation to a modified peripheral frequency map. *Acta Otolaryngol*, *124*, 381–386.

Svirsky, M. A., Talavage, T. M., Sinha, S., Neuburger, H., Azadpour, M. (2015b). Gradual adaptation to auditory frequency mismatch. *Hear Res*, *322*, 163–170.

Wilson, B. S., & Dorman, M. F. (2008). Interfacing sensors with the nervous system: Lessons from the development and success of the cochlear implant. *IEEE Sens J*, *8*, 131–147.

Winn, M. B., Chatterjee, M., Idsardi, W. J. (2021). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. *J Acoust Soc Am*, *131*, 1465–1479.

Winn, M. B., Rhone, A. E., Chatterjee, M., Idsardi, W. J. (2013). The use of auditory and visual context in speech perception by listeners with normal hearing and listeners with cochlear implants. *Front Psychol*, *4*, 824.

Winn, M. B., & Litovsky, R. Y. (2015). Using speech sounds to test functional spectral resolution in listeners with cochlear implants. *J Acoust Soc Am*, *137*, 1430–1442.

Winn, M. B., Won, J. H., Moon, I. J. (2016). Assessment of spectral and temporal resolution in cochlear implant users using psychoacoustic discrimination and speech cue categorization. *Ear Hear*, *37*, e377–e390.

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

Zhou, N., Xu, L., Lee, C. Y. (2010). The effects of frequency-place shift on consonant confusion in cochlear implant simulations. *J Acoust Soc Am*, *128*, 401–409.

REFERENCE NOTE

1. Goldstein, U. G. (1980). An Articulatory Model for the Vocal Tracts of Growing Children [doctoral dissertation]. Massachusetts Institute of Technology.