

Lexical bias in word recognition by cochlear implant listeners

Steven P. Gianakas and Matthew B. Winn^{a)}

Department of Speech-Language-Hearing Sciences, University of Minnesota, 164 Pillsbury Drive SE, Minneapolis, Minnesota 55455, USA

(Received 31 May 2019; revised 4 October 2019; accepted 14 October 2019; published online 18 November 2019)

When hearing an ambiguous speech sound, listeners show a tendency to perceive it as a phoneme that would complete a real word, rather than completing a nonsense/fake word. For example, a sound that could be heard as either /b/ or /g/ is perceived as /b/ when followed by _ack but perceived as /g/ when followed by “_ap.” Because the target sound is acoustically identical across both environments, this effect demonstrates the influence of top-down lexical processing in speech perception. Degradations in the auditory signal were hypothesized to render speech stimuli more ambiguous, and therefore promote increased lexical bias. Stimuli included three speech continua that varied by spectral cues of varying speeds, including stop formant transitions (fast), fricative spectra (medium), and vowel formants (slow). Stimuli were presented to listeners with cochlear implants (CIs), and also to listeners with normal hearing with clear spectral quality, or with varying amounts of spectral degradation using a noise vocoder. Results indicated an increased lexical bias effect with degraded speech and for CI listeners, for whom the effect size was related to segment duration. This method can probe an individual’s reliance on top-down processing even at the level of simple lexical/phonetic perception. © 2019 Acoustical Society of America.

<https://doi.org/10.1121/1.5132938>

[JFL]

Pages: 3373–3383

I. INTRODUCTION

Good speech intelligibility requires integration of the auditory signal (bottom-up cues) and our prior knowledge and expectations shaped by awareness of the situation and of the language (top-down cues). When the speech signal clarity is compromised, as for listeners with hearing loss or cochlear implants (CIs), the balance of importance for bottom-up and top-down cues may skew differently than for someone with typical hearing. The use of semantic context is an example of top-down processing that enables listeners to correctly perceive words within a sentence with more accuracy (Bilger *et al.*, 1984) and with less effort (Winn *et al.*, 2016). For example, when hearing “The baby drank from its ...,” the word “bottle” is easily perceived due to the high context information of the sentence. Listeners with hearing loss tend to rely more heavily on semantic context than listeners with normal hearing (NH) (Bilger *et al.*, 1984; Pichora-Fuller *et al.*, 1995), exhibiting the influence of top-down processing in the perception of speech when bottom-up cues are compromised. In standard clinical tests, words are presented in isolation, which removes many of these contextual cues. However, the influence of context not only occurs at sentence level, but can also impact perception at the word and phoneme level. In this study, we explore the influence of lexical status on phoneme perception, and how that influence is moderated by auditory spectral resolution and the use of a CI.

For the purpose of this study, lexical status simply refers to whether an utterance is a real word or not a real word in

the listener’s native language. Lexical status affects auditory recognition of words, especially when the interpretation of an ambiguous sound within that word hinges on whether it would mean the difference between hearing one phoneme or another. For example, a sound that could be heard as either /s/ or /ʃ/ is perceived as /s/ when preceded by “Christma_” but perceived as /ʃ/ when preceded by “fini_” (Norris *et al.*, 2003). In studies that examine word recognition in a full-sentence environment, consistency with preceding semantic context has an influence on how listeners identify the target phoneme in the sentence-final word (Borsky *et al.*, 1998; Connine, 1987). For example, a word that is phonetically ambiguous between *gap* and *cap* is more likely to be perceived as *gap* in “The girl jumped over the ___,” but as *cap* in “The boy put on the ___.” That effect appears to be further mediated by familiarity with talker accent and with exposure to testing stimuli (Schertz and Hawthorne, 2018). In addition, Vitevitch *et al.* (1997) showed that linguistic factors can influence perception at a sub-lexical level; the occurrence of a phoneme or the transitional probability between two phonemes also affects our perception of segments which are otherwise acoustically identical (e.g., at the beginning of a word, /s/ is more likely to be followed by /l/ than /f/). Documenting the relative co-occurrence of phonemes within words in isolation and within sentences, Boothroyd and Nittrouer (1988) established a mathematical framework for computing the influence of context on the likelihood of perceiving phonemes, in contrast to frameworks where phonemes are treated as independent units. The types of constraints computed by Boothroyd and Nittrouer (1988) are relevant to the current study because not all co-occurrences are equally likely, just as not all lexical items are equally frequent.

^{a)}Electronic mail: mwinn@umn.edu

An arguably more detailed perspective on the lexical bias effect comes not from whole-word decisions (as measured by Norris and other authors mentioned in the previous paragraph), but by examining effects on specific acoustic-phonetic boundaries. A phoneme's syllable environment might not completely change a perceptual judgment, but it could push it slightly one way or another. A classic example of the lexicon's influence on phoneme perception was done by Ganong (1980), whose experiment showed that lexical knowledge shifts phonetic boundaries in subtle but measurable ways consistent with a bias toward hearing real words. Exact acoustic-phonetic boundaries were measured by constructing a continuum of sounds that gradually change from one to the other on some dimension (like voice onset time, a formant transition, or whatever else distinguishes a phonetic contrast). This creates a psychometric function that reveals an estimation of the acoustic level where a listener's categorization crosses a 50% threshold, signaling a perceptual shift from one phoneme to the other. Ganong (1980) created a continuum between /g/ and /k/, which differ primarily by voice onset time (VOT; Lisker and Abramson, 1964). Importantly, the /g/-/k/ continuum was pre-appended to the context of “_iss” or “_ift,” which, according to lexical bias, should influence the listener to favor /k/ and /g/, respectively. In the case of “_iss,” the listener is biased toward /k/ because “kiss” is a word but “giss” is not; listeners were more willing to perceive short-VOT (i.e., more “voiced” /g/-like sounds) as /k/. Conversely, when the same sounds were preappended to “_ift,” the opposite bias was observed, because “gift” is a word but “kift” is not. Results indicated that in the “_ift” context, perception of /g/ was tolerable even for sounds that would otherwise be called /k/ in other phonetic environments. This lexical bias has sometimes been called the “Ganong effect.”

A. CIs and lexical bias

In this study, we recognize how ambiguity is likely to be a larger problem in the case of a person wearing a CI, where sensitivity to clear acoustic cues for phonemes is likely to be impaired even in quiet environments. CIs are surgically implanted hearing devices for individuals with severe-to-profound hearing loss who participate in the oral communication community. The implant bypasses the mechanical components of the cochlea to electrically stimulate the auditory nerve to provide a sensation of hearing. Despite the success of the implant, CI listeners have a decreased ability to perceive sounds in comparison to their NH peers (Fu and Shannon, 1998), particularly because of distortion in the representation of the spectrum. These distortions are exacerbated by the small number of spectral processing channels (Fishman *et al.*, 1997), the electrical interaction between electrodes (Chatterjee and Shannon, 1998; Abbas *et al.*, 2004), and incomplete electrode insertion depth during surgical placement (Svirsky *et al.*, 2015). For these reasons, listeners with CIs experience a degraded signal with coarse spectral information, which poses difficulties for hearing signals like speech, which has rapid spectral changes. Unsurprisingly, decreased spectral resolution is directly related to errors in speech perception (Loizou and Poroy, 2001).

Individuals with CIs commonly encounter ambiguity among word choices that are cued spectrally (i.e., by differences in the frequency spectrum), such as those differing by place of articulation. There are predictable difficulties in perceiving specific phonetic contrasts that are distinguished primarily by spectral cues (Munson *et al.*, 2003). We therefore hypothesize that spectral contrasts would be perceived as ambiguous, causing a CI listener to recruit extra lexical bias. In contrast to prior studies of the Ganong effect that used stimuli that varied by temporal dimensions, the current study focuses on spectrally-cued place-of-articulation contrasts.

B. Imposing spectral degradation on listeners with NH

Vocoders can be used to systematically change the spectral resolution of an acoustic signal in a way that mimics some elements of CI processing. CIs and vocoders operate on the shared principle of sound envelope extraction within frequency bands. Vocoders can therefore be used to better understand listening with the degraded spectral resolution of CIs. Noise-band vocoders (Friesen *et al.*, 2001) or sinewave vocoders (Fu and Shannon, 2000) are both used to test NH to approximate the intelligibility obtained in higher-performing CI users.

When spectral resolution is degraded listeners with NH or CIs tend to change their reliance on certain acoustic cues in speech perception. Specifically, non-spectral cues (such as temporal cues) can compensate when spectral information is compromised as a result of a noise vocoder (Xu *et al.*, 2005), consistent with reliance upon cues that remain relatively intact. Furthermore, when NH listeners use a vocoded signal, they tend to alter their use of phonetic cues in a way that is mostly reflective of behavior demonstrated by real CI listeners. For example, when listening to spectrally degraded vowels, listeners decrease their use of frequency-based formant cues and increase use of durational or temporal cues (Winn *et al.*, 2012; Moberly *et al.*, 2016) or other global acoustic properties (Winn and Litovsky, 2015).

C. Summary and hypotheses

Our study explores how spectral degradation in the case of CIs might result not only in poorer performance on basic psychophysical tasks, but also in an increased reliance upon linguistic information such as the lexical biases discussed above. In terms of speech perception and lexical processing, the uncertainty resulting from spectral degradation may influence listeners to rely more on lexical and linguistic knowledge to overcome missing information in the speech signal. In focusing on spectral contrasts, we set out to explore the influence of ambiguity that is specifically linked to the limitations of CIs. We hypothesize that degraded spectral resolution will elicit increased reliance upon lexical processing in order to categorize ambiguous phonemes, and that this effect should be even stronger in the case of phonetic contrasts with faster, more subtle acoustic contrasts. The quality of speech stimuli has previously been observed to moderate the lexical bias effect; noise masking appears to cause listeners to rely even more heavily on lexical status (Burton and Blumstein, 1995). More recently, Smart and

Frisch (2015) examined lexical influences on the perception of /t/ and /k/ sounds, using an approach similar to the one taken in the current study. Although no published results are available of their study of NH listeners, archived records of the work suggest the hypothesis that spectral degradation via vocoding would elicit an elevated effect of lexical bias in phonetic categorization. In their work, they used vocoded signals and found increased lexical bias for the /t/-/k/ contrast (Frisch, 2018). Taken together, the results discussed here suggest that when there are potential phonetic ambiguities relating to auditory degradation, lexical knowledge should play an important role in biasing perception.

Our second hypothesis is that the effect of lexical bias will be mediated by the difficulty of the particular phonetic contrast tested. The current study will focus on three phonetic contrasts that all differed by spectral cues, but which have three different durations. Vowels tend to be longer in duration and the slow formant information is used by listeners to differentiate them. For example, /a/ and /æ/ are both low unrounded vowels that mainly differ by vowel advancement as /æ/ is a front vowel with a higher second formant frequency than /a/. Because the vowel contrast is rather large spectrally and because the segment has longer duration (~240 ms), we expect that listeners will perceive less ambiguity during vowel perception in comparison to other quicker phonetic contrasts. The medium-speed contrast was /s/-/ʃ/ (~170 ms frication duration); these are both voiceless fricatives that differ in their spectral peak frequency with /ʃ/ having a lower spectral peak (Evers *et al.*, 1998; McMurray and Jongman, 2011). As for the slow vowel contrast, we expect that the acoustic-phonetic cue will be easily discriminable (based on /s/-/ʃ/ categorization functions in CI listeners obtained by Winn *et al.*, 2013), but that the relatively shorter exposure to the segment might yield slightly more difficulty than for the vowels. The fast contrast was between /b/ and /g/ (~85 ms duration of formant transition), which are both voiced plosives that differ by place of articulation. /b/ is a bilabial consonant that has formants that transition upwards in frequency while /g/ is a velar in which *F*₁ transitions upward, but where *F*₂ and *F*₃ begin pinched together and then diverge (Johnson, 2003). The relatively subtle frequency differences and quick formant transition duration of these two phonemes would be highly compromised with degraded spectral resolution, and we thus expect the most influence of lexical bias for this particular phonetic contrast.

II. METHODS

A. Participants

Participants included 20 adults between the ages of 18 and 39 years old with NH, defined as having pure-tone thresholds ≤ 20 dB HL from 250 to 8000 Hz in both ears (ANSI, 2004). A second group of participants included 21 adults (age 23–87; mean age 63 years, median age 64 years) listeners with CIs; see Table I for demographic information. All but two of the CI listeners were post-lingually deafened. Two were bilateral users who were tested separately in each ear. All listeners were fluent in North American English. There was a substantial age difference between CI and NH listener

TABLE I. Demographic information for CI participants in this study. Note: N5, N6, and N7 were processors for an array manufactured by the Cochlear Corporation; C1 Harmony, CII Harmony, and Naida are manufactured by Advanced Bionics, the Sonnet device is manufactured by MED-EL.

Listener	Sex	Age	Device	Ear	Etiology	CI Exp. (years)
C101	F	54	Sonnet	Bilateral	Sudden SNHL	5
C102	F	64	N6	Right	Idiopathic	2
C104	M	64	C1 Harmony	Bilateral	Ototoxicity	15
C105	F	47	N6	Bilateral	Progressive SNHL	8
C106	M	87	Naida Q90	Bilateral	Noise-related SNHL	30
C107	M	68	Unknown	Bilateral	Unknown	6
C110	M	78	N6	Bilateral	Progressive SNHL	14
C114	M	72	Naida Q90	Bilateral	Progressive SNHL	10
C116	F	61	Naida Q70	Right	Rheumatic fever	22
C117	M	66	Naida Q70	Bilateral	Auditory neuropathy	7
C119	F	22	N7	Bilateral	Unknown	17
C120	F	79	Sonnet	Left	Unknown	2
C121	M	52	N5	Right	Congenital, CP	22
C122	F	72	Naida Q70	Bilateral	Genetic	15
C123	F	60	CII Harmony	Left	Genetic	9
C124	M	62	Naida Q90	Right	Congenital	12
C128	F	23	N7	Left	Genetic	8
C129	M	81	N7	Right	Unknown	1
C131	F	69	N6	Right	Chronic ear infections	4
C132	M	80	N6	Right	Otosclerosis	4
C138	F	60	Naida Q90	Bilateral	Unknown	27

groups, mainly due to the availability of the test populations. Seven CI listeners were tested at the University of Washington, two at Stanford University, and 12 at the University of Minnesota. While the CI listeners utilized different devices with different signal coding strategies, they do share the common challenge of poor spectral resolution.

B. Stimuli

There were three speech continua that varied by spectral cues of varying speeds, including fast (stop formant transitions), medium (fricative spectra), and slow (vowel formants), described below. Speech stimuli consisted of modified natural speech tokens that were originally spoken by a native speaker of American English. Stimulus context were chosen to minimize any bias other than lexical bias (e.g., lexical frequency or familiarity). Real-word targets within each stimulus set were chosen with the intention of having roughly equivalent and sufficient lexical frequency and familiarity. Our initial verification of those features was based on the HML database, which is an online database of 20 000 words containing lexical frequency and familiarity ratings (Nusbaum *et al.*, 1984). The familiarity and frequency scores from the HML database were also comparable to the scores found on a more recent lexical neighborhood and phonotactic probability database (Vitevich and Luce, 2004). The HML database rates familiarity from 1 (least familiar) to 7 (most familiar) with a mean of 5.5. Any familiarity rating between 5.5 indicates that the word is familiar and the meaning is well known (Nusbaum *et al.*, 1984); all real-word stimuli had familiarity ratings of 7 except for “dash” (6.916). The frequency of words is quantified in

usage per 1×10^6 words of text, with a mean of 40.8. All real-word stimuli had similar frequency ratings; Dash was 11, “Dot,” was 13, “Safe” was 58, “Share” was 98, “Gap” was 17, and “back,” which was more common at 967.

For all stimuli, phonetic environments surrounding the segment of interest were kept consistent across stimulus set using a cross-fading/splicing method, described in detail for each stimulus below. All stimuli modifications were complete using the Praat software. Stimuli are available to download via the supplementary materials.¹

1. /æ/-/a/ continuum

A continuum from /æ/ to /a/ was created in the /d_ʃ/ context, creating sounds that ranged from dash (a real word) and “dosh” (a non-word), and also in the /d_t/ context (ranging from “dat” to dot). Recordings of the word dash and dot spoken by a native speaker of American English were manipulated to produce a continuum from /æ/ to /a/, using the method described by Winn and Litovsky (2015). In short, the vocalic portion of this syllable from the offset of the /d/ burst to the vowel/consonant boundary was decomposed into voice source and vocal tract filter using the LPC inverse-filtering algorithm in Praat. The filter was represented using FormantGrids that were systematically modified between formant contours for the original vowel (/æ/) and the formant contours for a recording of the vowel /a/ in dosh. Formants 1 through 4 were sampled at 30 time points throughout the vowel; continuum steps were interpolated using the Bark frequency scale to account for auditory non-linearities in frequency perception. High-frequency energy normally lost during LPC decomposition was restored using the method described in full detail by Winn and Litovsky (2015). A uniform /d/ burst (with prevoicing) was preappended to each vowel.

Using the continuum of /dæ/ to /da/, we created two lexical environments by preappending the continuum steps to either a 300-ms /ʃ/ segment excised from the original recording of dash or to the /t/ (including 94 ms closure gap) excised from the original recording of dot. The final result was a set of two continua: one that changed from dash to dosh (where lexical bias would favor the /æ/ endpoint) and another that changed from dat to dot (where bias should favor the /a/ endpoint). See Fig. 1 for an illustration of the stimulus construction for the dash-dosh and dat-dot continua.

2. /s/ vs /ʃ/ continuum

A continuum from /s/ to /ʃ/ was created and preappended to /_eɪf/ (“shafe,” a non-word, and “safe,” a real word) and /_eɪr/, (share, a real word and “sare,” a non-word). /s/ and /ʃ/ fricative segments of equal duration were extracted from recordings of safe and share. The fricative segments were subject to a blending/mixture procedure where the amplitude of one segment was multiplied by a factor intermediate between 0 and 1, and the other segment was multiplied by 1 minus that factor. For example, /s/ would be modified by 0.2 and /ʃ/ would be modified by 0.8. These fricatives were preappended to the vowel and offset of safe. The differences in vowel-onset formant transitions were

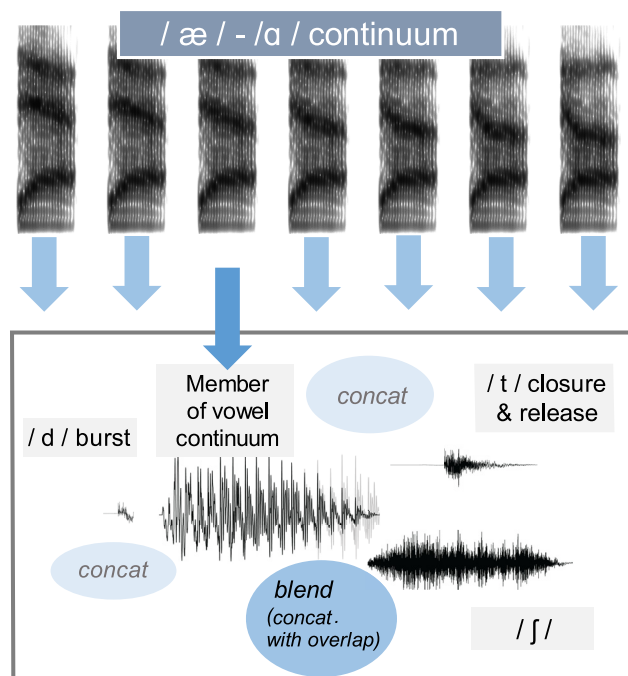


FIG. 1. (Color online) Illustration of stimulus construction for the dash-dosh and dat-dot continua. Concat refers to concatenation of two waveform segments. Blend refers to cross-fading of two waveform segments.

discarded based on the judgment by the authors that the /s/-onset vowel permitted natural-like perception of either fricative (furthermore, any residual effects of formant transitions would be balanced across opposing lexical contexts).

To create the /_eɪr/ environment, the first half of the /_eɪf/ syllable was spliced onto the second half of the syllable share so that regardless of phonetic environment (/_eɪf/ or /_eɪr/), the syllable onset (including fricative and vowel onset) remained exactly the same, removing any acoustic bias other than the segments that determined the word offset. Splicing was done using a cross-fading procedure where an 80-ms offset ramp of the first half was combined with an 80-ms onset ramp for the second half to ensure a smooth transition. The exact placement of the cross-splicing boundary was chosen strategically to avoid any discontinuities in envelope periodicity. The final result was a set of two continua: one that changed from safe to shafe (where bias should favor /s/) and another that changed from share to sare (where bias should favor /ʃ/). For the share- sare continua, duration of the initial fricative was roughly 220 ms, the /eɪ/ was approximately 162 ms (although there is not a straightforward transition into the /r/); the whole syllable coda was approximately 334 ms. The safe-shafe continua had the same durations except for the coda, /f/ which was approximately 321 ms. See Fig. 2 for an illustration of the share-sare and shafe-safe continua.

3. /b/ vs /g/ continuum

A continuum from /bæ/ to /gæ/ was created and preappended to /_p/ (with 85 ms closure duration) to make gap, a real word and “bap,” a non-word, or /_k/ (with 75 ms closure duration), to make “gack,” a non-word and back, a real word.

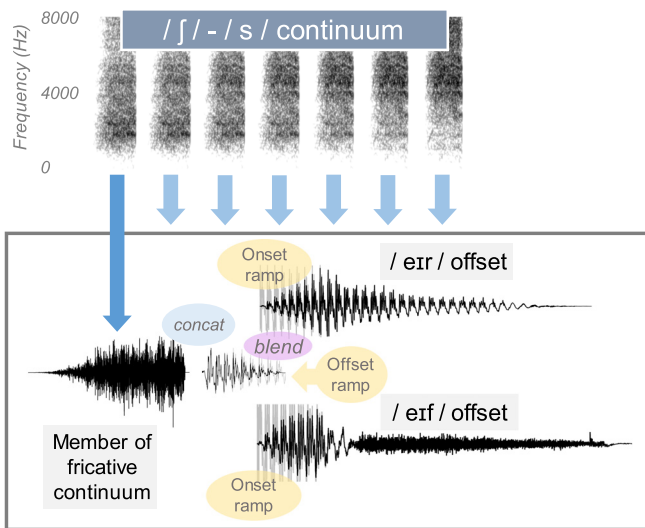


FIG. 2. (Color online) Illustration of the creation of the continuum for share-sare and shafe-safe. Concat refers to concatenation of two waveform segments. Onset and offset ramps were applied to blend the *e*/vowel onset into the offset portion of either the *_eɪr*/or *_eɪf*/vocalic portions, ensuring equivalent vowel onsets but natural transitions into the different syllable codas.

This continuum was made using the same basic procedure as for the /æ/ vs /ɑ/ continuum, where formant contours were imposed on a voice source that was derived from an inverse-filtered utterance. The onset prevoicing and release burst that was preappended to the vowel was a blend of 50% /b/ and 50% /g/ onsets so that the only cue available for the distinction was the set of formant transitions. After generating a continuum of bap-gap syllables, the /k/ from back was appended to the vocalic portion, which was truncated to remove glottalization that could cue /p/. This step was taken with a similar motivation as for the /f/-/s/ continuum, so that regardless of word-final phonetic environment, the syllable onset remained exactly the same within each continuum. The final result was a set of two continua: one that changed from back to gack [where bias should favor (/b/)] and another that changed from gap to bap (where bias should favor /g/)] (see Fig. 3).

C. Noise vocoding

To degrade spectral resolution, we used a conventional noise-channel vocoder (cf. Shannon *et al.*, 1995) implemented using Praat (Boersma and Weenink, 2016). Stimuli were band-pass filtered into 8 or 24 frequency bands to reflect poorer and better spectral resolution, respectively. The frequency spacing was designed so that each channel would occupy roughly equal cochlear space; specific band frequency cutoff values were determined assuming a 35 mm cochlear length using the function supplied by Greenwood (1990). The temporal envelope from each band was extracted by half-wave rectification and low-pass filtered with a 300-Hz cutoff frequency. The envelope of each band was used to modulate white noise that was then bandpass filtered to have the same frequency spectrum as its corresponding analysis filter. The channels were summed to create the final stimulus that sounded like a spectrally degraded version of the original word.

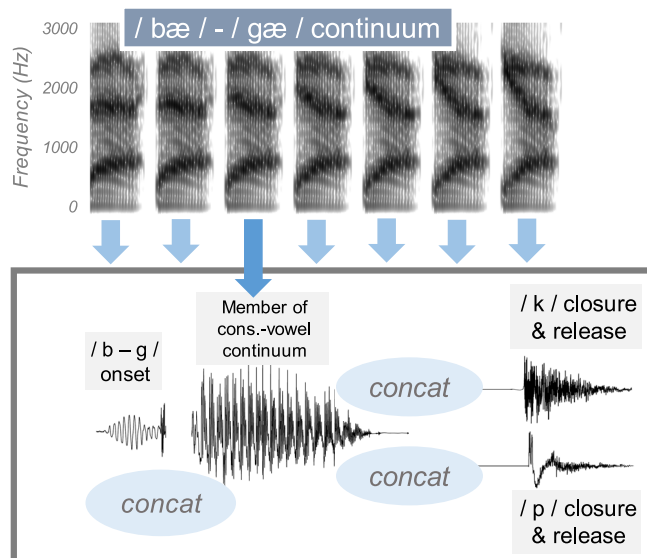


FIG. 3. (Color online) Illustration of the creation of the continuum for back-gap and gap-bap. Concat refers to concatenation of two waveform segments.

D. Procedure

Testing was conducted in a double-walled sound-treated booth. Stimuli were presented in quiet at 65 dBA in the free field through a single loudspeaker. A single stimulus was presented once, and listeners subsequently used a computer mouse to select one of four word choices (/æ/-/ɑ/ contrast: Dash, Dosh, Dat, or Dot), (/f/-/s/- contrast: Shafe, Safe, Share, or Sare), (/b/-/g/ contrast: Bap, Gap, Back, or Gack) to indicate their perception. Stimuli were presented in blocks organized by continuum and by degree of spectral resolution (unprocessed, noise vocoded 8- or 24-channel). Each unique stimulus was heard 6 times each; CI listeners listened only to the unprocessed speech through their everyday clinical processors using their normal parameters. Ordering of blocks was pseudo-randomized (for NH listeners, the first block was always one of the unprocessed speech blocks), and ordering of tokens within each block was randomized. Before testing, there were brief practice blocks with both normal speech (for all listeners) and 8-channel vocoded speech (for NH listeners). The testing conditions and protocol were equivalent across the three testing sites.

E. Analysis

For each continuum, we modeled the likelihood of listeners perceiving a specific end of the continuum (e.g., /ɑ/, /s/, or /g/) as a function of both the continuum step and the environment (e.g., *_æp* or *_æk*). Listeners' responses were fit using a generalized linear binomial (logistic) mixed-effects model (GLMM) using the lme4 package (Bates *et al.*, 2015) in the R software interface (R Core Team, 2016). The binomial family call function was utilized because responses were coded in a binary fashion (i.e., 0 or 1 re continuum endpoints).

Continuum step was coded to represent each listener's personalized category boundary, interpreted as the continuum step that produced the greatest influence of lexical bias.

For example, if a listener showed the greatest bias at step 4, then step 4 was coded as step 0, with steps 3 and 5 coded as -1 and $+1$, respectively. This personalized centering procedure was done separately for each listener, in each condition, for each speech contrast.

The main effects were continuum step (slope), continuum series ($\text{æ}-\text{a}$, $\text{s}-\text{ʃ}$, $\text{b}-\text{g}$), lexical environment (word/non-word), and listening condition (normal/24-channel vocoder/8-channel vocoder/CI). There were random effects of intercept, slope, condition, lexical bias effect, and the interaction between lexical bias and condition for each listener. For the purpose of simplicity in reporting, each phonetic contrast was modeled separately.

III. RESULTS

A. Overall results and lexical bias

Average group responses to the three phonetic contrasts are displayed in Fig. 4. All functions have a sigmoidal shape and spanned from nearly 0 to nearly 1, suggesting that perception of phoneme continuum endpoints was reliable in all conditions. As seen by the difference between lines in Fig. 4, lexical bias emerged reliably for nearly all conditions (Normal, and both vocoder conditions) for both NH and CI listeners. The direct magnitude of the lexical bias effect (i.e., difference between curves in Fig. 4) across the continuum for each contrast is displayed in Fig. 5 (upper panels). Visual inspection of these figures suggests that target phonemes were more likely to be heard as sounds that made a real word, rather than a non-word (consistent with Ganong, 1980), particularly when spectral resolution was degraded. For example, $/\text{æ}/$ was more likely to be perceived when the environment was such that $/\text{æ}/$ rendered the syllable a real word, like “d_sh.”

An unplanned follow-up analysis was performed on the CI listener data for the $/\text{b}/$ - $/\text{g}/$ contrast in light of observations that there were proportionally more errors in perceiving the syllable coda. This was an important factor to consider, since misperception of back as bap would essentially reverse the direction of the lexical bias, weakening the lexical effect due to auditory deficiency rather than true lexical-bias disposition. This follow-up analysis compared CI results for the full $/\text{b}/$ - $/\text{g}/$ dataset to data only where the lexical bias status was coded according to the perceived coda rather than the stimulus itself (i.e., responses ending in $_{\text{ack}}$ were recoded as if the stimulus were a member of the bap-gap continuum, and responses ending in $_{\text{ap}}$ were recoded as if the stimulus were a member of the back-gack continuum). The separate results are displayed in Fig. 6 and statistical analysis confirmed that the lexical bias effect was larger for the “correct-series-only” responses ($\beta = 0.254$, standard error = 0.139, $z = 1.831$), but did not reach the conventional criterion for statistical significance ($p = 0.067$). In other words, lexical bias tends in the direction of the perceived word category, even if the perception is incorrect.

The generalized linear mixed-effects (GLM) model summaries are shown in Table II, where beta estimates, z values, and p values are listed. Each specific model term is labeled for ease of reference throughout this text. The most important term in these models is the effect of real-word environment (A5), and how that changed across conditions (technically, the interaction of the lexical environment and the condition; A6 through A8). For the $/\text{æ}/$ - $/\text{a}/$ contrast, real-word environments favoring $/\text{a}/$ (i.e., “d_t”) significantly increased the likelihood of $/\text{a}/$ responses for the ideal NH condition (A5; $p = 0.023$). Though numerically larger than in the unprocessed condition, this effect was not statistically

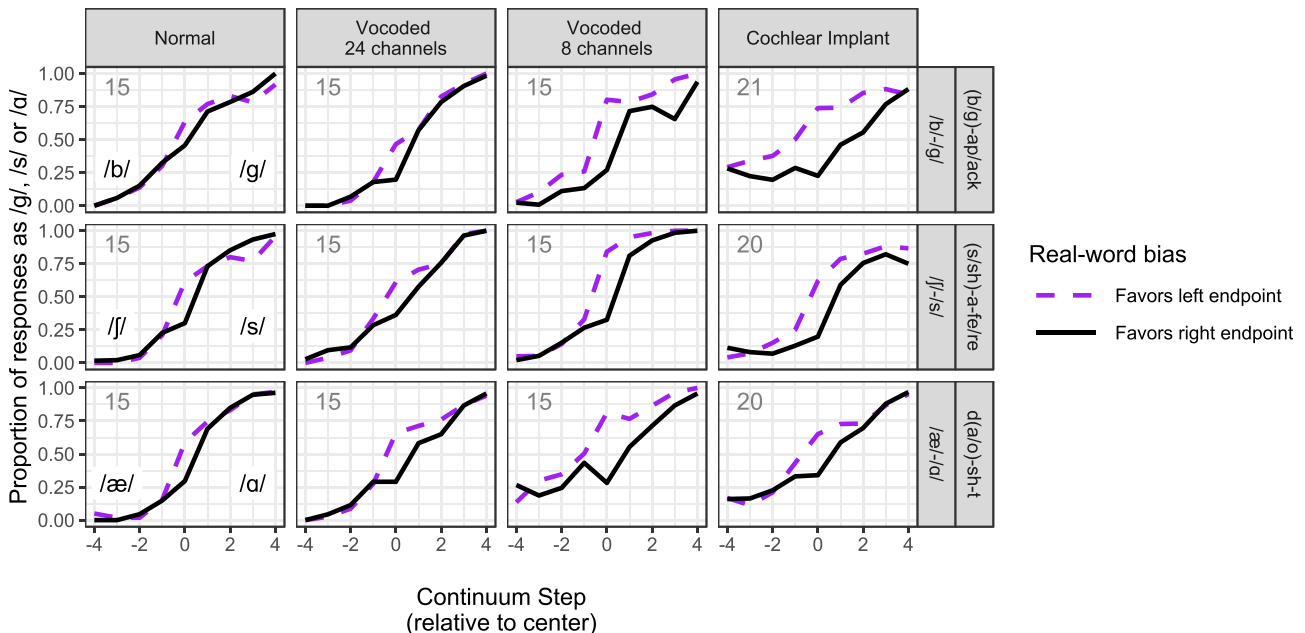


FIG. 4. (Color online) Proportion of phoneme perceived as $/\text{g}/$, $/\text{s}/$, or $/\text{a}/$ along the phoneme continua (rows) in the normal, vocoded, and CI conditions (columns). The dashed line indicates perception of the target phoneme in a real word environment that would favor as $/\text{g}/$, $/\text{s}/$, or $/\text{a}/$ perception, and the solid line indicates perception in a real word environment that would favor the opposing member of each continuum. Greater lexical bias is observed as more separation between the dashed line and the solid line. Inset numbers represent the number of listeners contributing data in each panel.

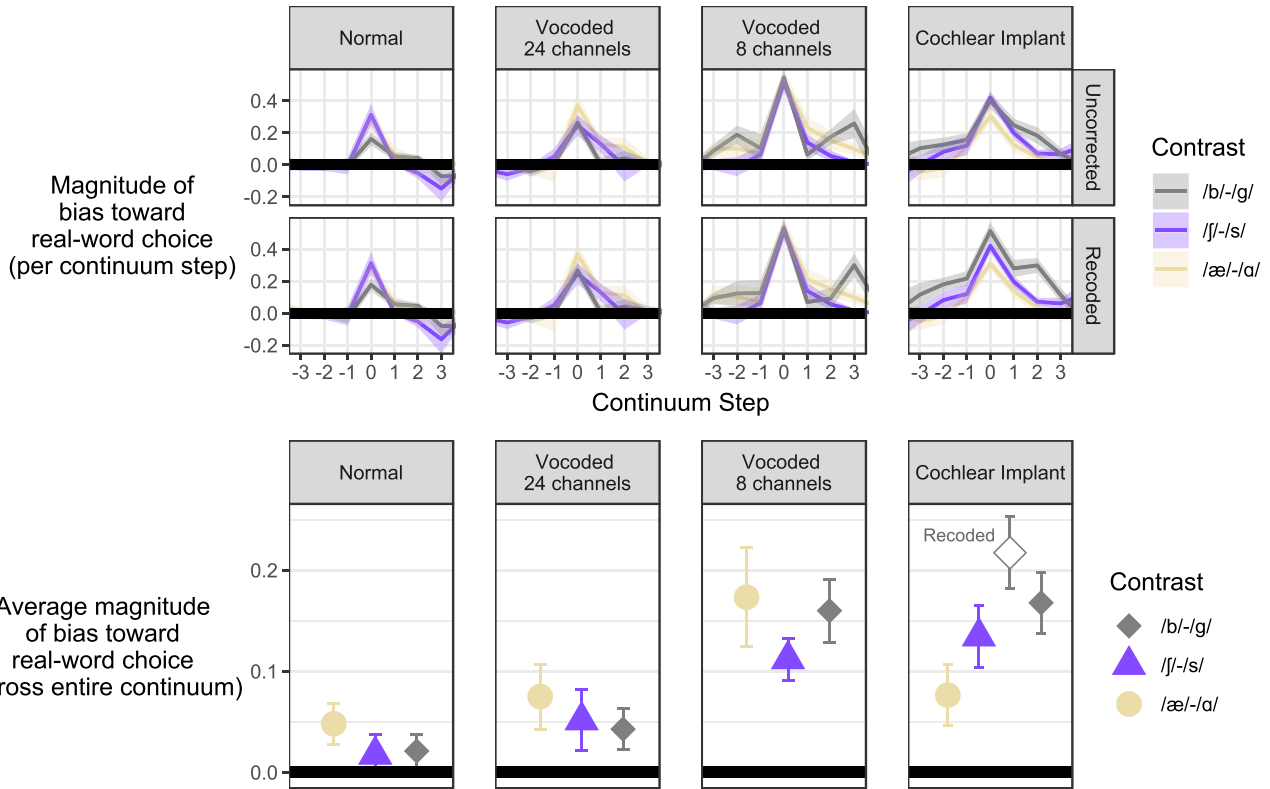


FIG. 5. (Color online) Magnitude of lexical bias across the phoneme continuum. Top row: The difference between lines from Fig. 4 is plotted across the continuum, where higher points indicate greater response bias in favor of real word perception. Data in the top row include mistakes across continua, meaning listeners may have responded with *_ack* for a *_ap* contextual environment. Middle row: Lexical bias across the continuum is plotted for corrected data in which lexical bias was coded to be consistent with whatever continuum the listener perceived (e.g., misperception of *_ack* as *_ap* was coded as promoting */g/* rather than */b/*). Bottom row: The average magnitude of lexical bias across the entire continuum expressed as a single point. For CI listeners, the filled diamond reflects the size of the effect for the full dataset that included perceptual errors on the continuum series. The open diamond is for responses that were recorded, as described above.

different in either of the vocoded conditions nor for the CI listeners (A6 through A8).

For the */f/-/s/* contrast, the lexical bias in the unprocessed condition (S5) was not detectable. The lexical bias effect in the 24-channel condition (S6), 8-channel condition (S7), and for CI listeners (S8), were all numerically larger than that in the unprocessed condition, but none reached statistical significance.

For the */b/-/g/* contrast, the lexical bias effect was not statistically detectable in the unprocessed condition (G5), but grew to reach significance in the 8-channel vocoder condition (G7) and for CI listeners (G8).

B. Function slopes/reliability of perceiving the phonetic continua

In addition to the main predicted effects of lexical bias, the model included a term for measuring the slope of each psychometric function and how it changed across conditions. We hypothesized that the effect of a change in step within the continuum would decrease in the degraded conditions, which would correspond to the shallower slope of psychometric functions for vocoded conditions and CI listeners as displayed in Fig. 4. This hypothesis was confirmed by the GLMM, where there were significant negative interactions between the continuum step term and each of the terms for condition (A10 through A12; S10 through 12; G10 through 12) (deviating from the NH condition; A9, S9, G9). For all three contrasts, there was the predicted

sequentially shallower slope for the NH, 24-channel, 8-channel and CI conditions. These results are consistent with previous work showing poorer phoneme categorization when spectral resolution is poor (Winn and Litovsky, 2015).

IV. DISCUSSION

When encountering an ambiguous phoneme, listeners were more likely to perceive it as a phoneme that would complete a real word rather than a non-word, replicating the classic Ganong effect. This effect was observed more

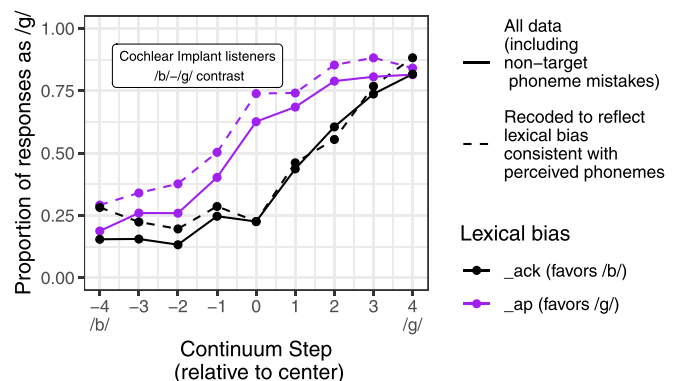


FIG. 6. (Color online) Re-plotting of psychometric functions for CI listeners for the */b/-/g/* contrast, separately grouping responses that were internally corrected (dashed lines) versus the full dataset, which included some errors across continuum series (solid lines).

TABLE II. Results of GLM. Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, $p < 0.1$.

/æ/ - /a/ Continuum (1 = /a/ response)		Estimate	St. Err.	z	Pr(> z)	
A1	Intercept (NH unprocessed)	-0.875	0.406	-2.155	0.031	*
A2	Vocoder 24 Ch.	-0.190	0.902	-0.210	0.833	
A3	Vocoder 8 Ch.	0.555	0.690	0.804	0.421	
A4	CI	0.260	0.542	0.479	0.632	
A5	/a/-makes-real-word (lexical bias)	0.557	0.245	2.277	0.023	*
A6	/a/-makes-real-word: Vocoder 24 Ch.	0.215	0.286	0.751	0.453	
A7	/a/-makes-real-word: Vocoder 8 Ch.	0.669	0.443	1.510	0.131	
A8	/a/-makes-real-word: CI	0.048	0.336	0.144	0.886	
A9	Continuum Step (slope)	1.797	0.138	13.069	<0.001	***
A10	Continuum Step: Vocoder 24 Ch.	-0.215	0.063	-3.442	<0.001	***
A11	Continuum Step: Vocoder 8 Ch.	-0.931	0.053	-17.451	<0.001	***
A12	Continuum Step: CI	-0.757	0.179	-4.227	<0.001	***
/ʃ/ - /s/ Continuum (1 = /s/ response)		Estimate	St. Err.	z	Pr(> z)	
S1	Intercept (NH unprocessed)	-0.974	0.769	-1.267	0.205	
S2	Vocoder 24 Ch.	2.632	0.200	13.172	<0.001	***
S3	Vocoder 8 Ch.	0.419	0.404	1.038	0.299	
S4	CI	-0.312	1.545	-0.202	0.840	
S5	/s/-makes-real-word (lexical bias)	0.938	0.866	1.083	0.279	
S6	/s/-makes-real-word: Vocoder 24 Ch.	-0.160	0.830	-0.193	0.847	
S7	/s/-makes-real-word: Vocoder 8 Ch.	0.387	0.449	0.861	0.389	
S8	/s/-makes-real-word: CI	0.632	0.499	1.267	0.205	
S9	Continuum Step (slope)	1.051	0.486	2.161	0.031	*
S10	Continuum Step: Vocoder 24 Ch.	-0.455	0.108	-4.207	<0.001	***
S11	Continuum Step: Vocoder 8 Ch.	-1.065	0.097	-10.944	<0.001	***
S12	Continuum Step: CI	-1.093	0.257	-4.259	<0.001	***
/b/ - /g/ Continuum (1 = /g/ response)		Estimate	St. Err.	z	Pr(> z)	
G1	Intercept (NH unprocessed)	-0.716	0.887	-0.807	0.420	
G2	Vocoder 24 Ch.	-0.736	0.945	-0.779	0.436	
G3	Vocoder 8 Ch.	0.005	0.918	0.005	0.996	
G4	CI	-0.187	0.910	-0.206	0.837	
G5	/g/-makes-real-word (lexical bias)	0.269	0.237	1.136	0.256	
G6	/g/-makes-real-word: Vocoder 24 Ch.	0.294	0.357	0.823	0.410	
G7	/g/-makes-real-word: Vocoder 8 Ch.	1.425	0.474	3.007	0.003	**
G8	/g/-makes-real-word: CI	1.036	0.317	3.270	0.001	**
G9	Continuum Step (slope)	1.791	0.153	11.684	<0.001	***
G10	Continuum Step: Vocoder 24 Ch.	-0.028	0.063	-0.443	0.658	
G11	Continuum Step: Vocoder 8 Ch.	-0.399	0.056	-7.125	<0.001	***
G12	Continuum Step: CI	-1.059	0.196	-5.400	<0.001	***

strongly when the speech was ambiguous (i.e., when a stimulus was in the middle of a phonetic continuum), especially when the signals were degraded. The effect of degradation to increase lexical bias was not as strong as expected, but did trend in the hypothesized direction for all contrasts. This effect extended to listeners who use CIs, consistent with what is known about the spectral degradation experienced by this population. The increase in magnitude of lexical bias in spectrally degraded conditions in this study supports the hypothesis that signal ambiguity promotes increased reliance on non-acoustic information. Furthermore, when controlling for correct perception of lexical bias environments (for /b/-/g/ stimuli), CI listeners showed increased lexical bias with faster phonetic contrasts (i.e., more lexical bias for /ʃ/-/s/ than for /æ/-/a/ and most lexical bias for /b/-/g/), suggesting that faster contrasts may have increased difficulty and thus demand more lexical influence.

Our findings are consistent with models of perception that explicitly incorporate lexical processing. For example, in the TRACE model (McClelland and Elman, 1986), speech perception involves the integration of phonemic processing, details of the speech input, and lexical processes. For a phoneme that is ambiguous between /b/ and /g/ in the contextual environment of _ack, gradual buildup of phoneme activation occurs followed by word-level activation once more of the speech signal is perceived. It is not until after the /k/ is introduced that the TRACE model obtains the information necessary to determine which lexical choice is closest to the input, (i.e., closer to back, rather than gack, or other possible activations such as “bat”). More activation is likely for the lexically plausible word, leading to a bias consistent with those observed in this study. Hence, it has long been thought that higher-level interpretative and lower-level acoustic processes interact to resolve ambiguities in a signal. The inclusion of

this effect in formal theories demonstrates a fundamental principle: perception of speech is influenced, not driven merely by acoustics but also by the structure and statistics of the language. Furthermore, this influence is gradient rather than categorical, as suggested by the different strengths of the effect in the studies described above.

Listeners experiencing the reduced spectral resolution of a CI tend to weight cues differently and rely *more* heavily on top-down processes compared to their peers with typical clear acoustic hearing. Our results are consistent with previous studies of increased reliance by CI listeners on non-auditory factors like visual cues (Desai *et al.*, 2008; Winn *et al.*, 2013) and semantic context (Loebach *et al.*, 2010; Patro and Mendel, 2016; Winn *et al.*, 2016). It can be hypothesized that the current results would generalize to listeners with hearing loss who do not use CIs, as individuals with cochlear hearing loss have poorer spectral resolution due to broadened auditory filters (Glasberg and Moore, 1986). It is not yet fully understood whether the extra reliance on non-auditory perception results in further downstream consequences, but other studies indirectly support this notion, since degraded auditory input also leads to increased listening effort (Winn *et al.*, 2015), slower word recognition (Grieco-Calub *et al.*, 2009), impaired and delayed lexical processing (Farris-Trimble *et al.*, 2014; McMurray *et al.*, 2017), slower processing of contextual information (Winn *et al.*, 2016), and poorer recognition memory (Van Engen *et al.*, 2012; Gilbert *et al.*, 2014).

In addition to degradation of spectral peaks in the signal, the use of a CI or vocoder would also likely compromise the encoding of dynamic spectral cues as well. This would likely have the greatest impact on the /b/-/g/ contrast, which is cued by a formant transition lasting only 85 ms or less. Formant dynamics could also play a large role in English vowel perception in general, with classic work demonstrating that formant onsets and offsets could permit excellent accuracy in vowel recognition even when the center of the vowel is omitted (Jenkins *et al.*, 1983; Strange *et al.*, 1983). However, this well-known pattern in NH listeners does not appear to fully generalize to CI listeners, who do not demonstrate the same sensitivity to dynamic spectral cues when presented with silent-center vowels (Donaldson *et al.*, 2015).

There was an age difference between the CI and NH hearing listener groups in the current study, which posed a complication for the interpretation of the results. Older adults tend to rely more on top-down influences than younger adults, although this may be attributed to a reduced ability to suppress lexical competition (Mattys and Scharenborg, 2014) and even more so when cognitive load is increased (Mattys and Wiget, 2011). Therefore the effects measured in CI listeners could have been a mixture of poor spectral resolution (substantiated by the vocoder results) as well as an effect of aging.

It is important to note that vocoders do not perfectly mimic hearing with CIs, in several ways that could be relevant to the topic under study here. Conventional channel vocoders typically do not convey spread of excitation and other factors related to the transmission of the electrical signal, such as pulsatile stimulation, channel peak-picking,

increased neural phase locking, and compressed dynamic range. There are exceptions to these trends (Deeks and Carlyon, 2004; Litvak *et al.*, 2007; Bingabr *et al.*, 2008; Stafford *et al.*, 2013; Winn *et al.*, 2015). As an alternative to sinusoidal vocoders, this current study utilizes simple conventional noise-band vocoders to explore the influence of spectral degradation in a way that should help further understand the experience of people with CIs, though not the total experience of hearing with these devices. A noise vocoder was used because it removes harmonicity and smears spectral detail such as formants, which are important for the perception of the selected phonemes in this study. Additionally, noise vocoders remove voice pitch cues from the speech signal, like with the use of a CI. In contrast, voice pitch can remain intact with the use of a sinusoidal vocoder, especially with the use of high envelope cutoff frequencies, which is not seen in typical performance of CI listeners (Souza and Rosen, 2009). Moreover, the phonemes contrasts selected in this study mainly differed in formant information or spectral peak, which would be noticeably degraded by the noise vocoder's ability to smear spectral information. Despite lack of a perfect match, the vocoder can be understood to be a tool to easily degrade the spectral resolution of an auditory signal, regardless of its status as a proper CI "simulation."

Rather than being merely an oddity of laboratory behavior, the lexical bias effect is likely advantageous in real-life situations where phonetic detail can be compromised by background noise, hearing loss, or misarticulation. The current experiment might serve as a clinical tool for revealing a patient's inability to rely solely on auditory cues. Importantly, the lexical bias demonstrated in this study would not be detected on standard clinical word recognition tests that consist of all real words (i.e., all "endpoint" stimuli). In the case that an individual misperceives the word gap as bap, they might respond *correctly* with gap, which would award credit in the case of a mistaken auditory perception corrected by lexical knowledge. With the standard battery of clinical tests, it cannot be known how often this type of mistake-correction process occurs.

V. CONCLUSIONS

Lexical knowledge influences phonetic perception when phonemes are ambiguous, not only because of phonetically ambiguous acoustic cues but also due to poor spectral resolution. Poor spectral resolution created by vocoders and CIs can make speech more ambiguous to a listener. As spectral resolution is diminished further, listeners become more likely to rely on lexical bias, although not equally for all phonetic contrasts. Listeners with CIs showed greater lexical bias for shorter/faster phonemes. Individuals with hearing loss might produce correct responses in clinical tests that reflect lexical effects that masquerade as successful auditory processing.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. R03 DC 014309 to M.B.W. and the American Speech-Language-Hearing Association Students Preparing for Academic and

Research Careers Award to S.P.G. We are grateful to Ashley N. Moore, David J. Audet, Jr., Tiffany Mitchell, and Josephine A. Lyou for their assistance with data collection and equipment setup. Stefan Frisch provided extra insight into the significance of this project. In addition, we are grateful to Matthew Fitzgerald, Ph.D. and the audiology team at Stanford Ear Institute as well as Kate Teece for their help in participant recruitment. Portions of this paper were presented as a poster at the 2016 fall meeting of the Acoustical Society of America (Honolulu, HI), AudiologyNOW! 2017 (Indianapolis, IN), and the Conference on Implantable Auditory Prostheses 2017 (Lake Tahoe, CA).

¹See supplementary material at <https://doi.org/10.1121/1.5132938> for stimulus audio files and corresponding TextGrids.

- 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. Neurotol.* **9**, 203–213.
- ANSI S3.6-2004. (2004). "American National Standard Specification for audiometers" (American National Standards Institute, New York).
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). "Fitting linear mixed-effects models using lme4." *J. Stat. Software* **67**(1), 1–48.
- Bilger, R., Nuetzel, J., Rabinowitz, W., and Rzeczkowski, C. (1984). "Standardization of a test of speech perception in noise." *J. Speech Hear. Res.* **27**(March), 32–48.
- Bingabr, M., Espinoza-Varas, B., and Loizou, P. (2008). "Simulating the effect of spread of excitation in cochlear implants." *Hear. Res.* **241**, 73–79.
- Boersma, P., and Weenink, D. (2016). "Praat: Doing phonetics by computer (version 6.0.21) [computer program]." <http://www.praat.org/> (Last viewed November 6, 2019).
- Boothroyd, A., and Nittrouer, S. (1988). "Mathematical treatment of context effects in phoneme and word recognition." *J. Acoust. Soc. Am.* **84**(1), 101–114.
- Borsky, S., Tuller, B., and Shapiro, L. (1998). "'How to milk a coat.' The effects of semantic and acoustic information on phoneme categorization." *J. Acoust. Soc. Am.* **103**, 2670–2676.
- Burton, M., and Blumstein, S. E. (1995). "Lexical effects on phonetic categorization: The role of stimulus naturalness and stimulus quality." *J. Exp. Psych.: Hum. Percept. Perf.* **21**, 1230–1235.
- Chatterjee, M., and Shannon, R. (1998). "Forward masked excitation patterns in multielectrode electrical stimulation." *J. Acoust. Soc. Am.* **103**, 2565–2572.
- Connine, C. (1987). "Constraints on interactive processes in auditory word recognition: The role of sentence context." *J. Memory Lang.* **16**, 527–538.
- Deeks, J. M., and Carlyon, R. P. (2004). "Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation." *J. Acoust. Soc. Am.* **115**, 1736–1746.
- Desai, S., Stickney, G., and Zeng, F. (2008). "Auditory-visual speech perception in normal-hearing and cochlear-implant listeners." *J. Acoust. Soc. Am.* **123**(1), 428–440.
- Donaldson, G. S., Rogers, C. L., Johnson, L. B., and Oh, S. H. (2015). "Vowel identification by cochlear implant users: Contributions of duration cues and dynamic spectral cues." *J. Acoust. Soc. Am.* **138**, 65–73.
- Evers, V., Reetz, H., and Lahiri, A. (1998). "Crosslinguistic acoustic categorization of sibilants independent of phonological status." *J. Phonetics* **26**, 345–370.
- Farris-Trimble, A., McMurray, B., Cigrand, N., and Tomblin, J. B. (2014). "The process of spoken word recognition in the face of signal degradation." *J. Exp. Psychol. Hum. Percept. Perform.* **40**(1), 308–327.
- Fishman, K., Shannon, R., and Slattery, W. (1997). "Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor." *J. Speech Lang. Hear. Res.* **40**, 1201–1215.
- Friesen, L., Shannon, R., Baskent, D., and Wang, X. (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants." *J. Acoust. Soc. Am.* **110**, 1150–1163.
- Frisch, S. (2018). (personal communication).
- Fu, Q., and Shannon, R. (1998). "Effects of amplitude nonlinearity on phone recognition by cochlear implant users and normal-hearing listeners." *J. Acoust. Soc. Am.* **104**, 2570–2577.
- Fu, Q., and Shannon, R. (2000). "Effect of stimulation rate on phoneme recognition by nucleus-22 cochlear implant listeners." *J. Acoust. Soc. Am.* **107**, 589–597.
- Ganong, W. F. (1980). "Phonetic categorization in auditory word perception." *J. Exp. Psych.: Hum. Percept. Perf.* **6**(1), 110–115.
- Gilbert, R., Chandrasekaran, B., and Smiljanic, R. (2014). "Recognition memory in noise for speech of varying intelligibility." *J. Acoust. Soc. Am.* **135**, 389–399.
- Glasberg, B. R., and Moore, B. C. (1986). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments." *J. Acoust. Soc. Am.* **79**, 1020–1033.
- Greenwood, D. (1990). "A cochlear frequency-position for several species—29 years later." *J. Acoust. Soc. Am.* **87**(6), 2592–2605.
- Grieco-Calub, T., Saffran, J., and Litovsky, R. (2009). "Spoken word recognition in toddlers who use cochlear implants." *J. Speech Lang. Hear. Res.* **52**(6), 1390–1400.
- Jenkins, J., Strange, W., and Edman, T. (1983). "Identification of vowels in 'vowelless' syllables." *Percept. Psychophys.* **34**, 441–450.
- Johnson, K. (2003). *Acoustic & Auditory Phonetics*, 2nd ed. (Blackwell, Malden, MA).
- Lisker, L., and Abramson, A. (1964). "A cross-language study of voicing in stops: Acoustical measurements." *Word* **20**, 384–422.
- Litvak, L., Spahr, A., Saoji A., and Fridman, G. (2007). "Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners." *J. Acoust. Soc. Am.* **122**, 982–991.
- Loebach, J. L., Pisoni, D. B., and Svirsky, M. A. (2010). "Effects of semantic context and feedback on perceptual learning of speech processed through an acoustic simulation of a cochlear implant." *J. Exp. Psych.: Hum. Percept. Perf.* **36**(1), 224–234.
- Loizou, P., and Poroy, O. (2001). "Minimum spectral contrast needed for vowel identification by normal hearing and cochlear implant listeners." *J. Acoust. Soc. Am.* **110**, 1619–1627.
- Mattys, S. L., and Scharenborg, O. (2014). "Phoneme categorization and discrimination in younger and older adults: A comparative analysis of perceptual, lexical, and attentional factors." *Psychol. Aging* **29**(1), 150–162.
- Mattys, S. L., and Wiget, L. (2011). "Effects of cognitive load on speech recognition." *J. Memory Lang.* **65**(2), 145–160.
- McClelland, J., and Elman, J. (1986). "The TRACE model of speech perception." *Cog. Psych.* **18**, 1–86.
- McMurray, B., Farris-Trimble, A., and Rigler, H. (2017). "Waiting for lexical access: Cochlear implants or severely degraded input lead listeners to process speech less incrementally." *Cognition* **169**, 147–164.
- 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." *Psychol. Rev.* **118**(2), 219–246.
- Moberly, A., Lowenstein, J., and Nittrouer, S. (2016). "Word recognition variability with cochlear implants: 'perceptual attention' versus 'auditory sensitivity.'" *Ear Hear.* **37**, 14–26.
- Munson, B., Donaldson, G., Allen, S., Collison, E., and Nelson, D. (2003). "Patterns of phoneme perception errors by listeners with cochlear implants as a function of overall speech perception ability." *J. Acoust. Soc. Am.* **113**(2), 925–935.
- Norris, D., McQueen, J., and Cutler, A. (2003). "Perceptual learning in speech." *Cognitive Psychol.* **47**, 204–238.
- 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 10, Speech Research Laboratory, Indiana University, Bloomington, IN, pp. 357–376.
- Patro, C., and Mendel, L. L. (2016). "Role of contextual cues on the perception of spectrally reduced interrupted speech." *J. Acoust. Soc. Am.* **140**, 1336–1345.
- Pichora-Fuller, K., Schneider, B., and Daneman, M. (1995). "How young and old adults listen to and remember speech in noise." *J. Acoust. Soc. Am.* **97**, 593–608.
- R Core Team (2016). "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria, www.R-project.org (Last viewed November 6, 2019).
- Schertz, J., and Hawthorne, K. (2018). "The effect of sentential context on phonetic categorization is modulated by talker accent and exposure." *J. Acoust. Soc. Am.* **143**, EL231–EL236.

- Shannon, R., Zeng, F., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues." *Science* **270**, 303–304.
- Smart, J., and Frisch, S. (2015). "Top-down influences in perception with spectrally degraded resynthesized natural speech." *J. Acoust. Soc. Am.* **138**, 1810.
- Souza, P., and Rosen, S. (2009). "Effects of envelope bandwidth on the intelligibility of sine- and noise vocoded speech." *J. Acoust. Soc. Am.* **126**(2), 792–805.
- Stafford, R., Stafford, J., Wells, J., Loizou, P., and Keller, M. (2013). "Vocoder simulations of highly focused cochlear implant stimulation with limited dynamic range and discriminable step." *Ear Hear.* **35**, 262–270.
- Strange, W., Jenkins, J., and Johnson, T. (1983). "Dynamic specification of coarticulated vowels." *J. Acoust. Soc. Am.* **74**, 695–705.
- Svirsky, M., Fitzgerald, M., Sagi, E., and Glassman, E. (2015). "Bilateral cochlear implants with large asymmetries in electrode insertion depth: Implications for the study of auditory plasticity." *Acta Oto-Laryngologica* **135**(4), 354–363.
- Van Engen, K., Chandrasekaran, B., and Smiljanic, R. (2012). "Effects of speech clarity on recognition memory for spoken sentences." *PLoS One* **7**(9), e43753.
- Vitevitch, M., Luce, P. A., Charles-Luce, J., and Kemmerer, D. (1997). "Phonotactics and syllable stress: Implications for the processing of spoken nonsense words." *Lang. Speech* **40**, 47–62.
- Vitevitch, M. S., and Luce, P. A. (2004). "A web-based interface to calculate phonotactic probability for words and nonwords in English." *Behav. Res. Methods, Instrum., Comput.* **36**, 481–487.
- Winn, M., Chatterjee, M., and Idsardi, W. J. (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., Edwards, J. R., and Litovsky, R. Y. (2015). "The impact of auditory spectral resolution in listeners with cochlear implants." *Ear Hear.* **36**(4), e153–e165.
- Winn, M., Rhone, A., Chatterjee, M., and Idsardi, W. (2013). "Auditory and visual context effects in phonetic perception by normal-hearing listeners and listeners with cochlear implants." *Front. Psych. Auditory Cogn. Neurosci.* **4**, 1–13.
- Winn, M. B., and Litovsky, R. Y. (2015). "Using speech sounds to test functional spectral resolution in listeners with cochlear implants." *J. Acoust. Soc. Am.* **137**(3), 1430–1442.
- Winn, M. B., Won, J. H., and 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., and Pfingst, B. (2005). "Relative contributions of spectral and temporal cues for phoneme recognition." *J. Acoust. Soc. Am.* **117**, 3255–3267.