Neural-scaled entropy predicts the effects of nonlinear frequency compression on speech perception

Varsha H. Rallapalli and Joshua M. Alexander[a)]
Department of Speech, Language and Hearing Sciences, Purdue University, 715 Clinic Drive, Lyles-Porter Hall, West Lafayette, Indiana 47907, USA

(Received 19 August 2014; revised 1 October 2015; accepted 15 October 2015; published online 17 November 2015)

The Neural-Scaled Entropy (NSE) model quantifies information in the speech signal that has been altered beyond simple gain adjustments by sensorineural hearing loss (SNHL) and various signal processing. An extension of Cochlear-Scaled Entropy (CSE) [Stilp, Kiefte, Alexander, and Kluender (**2010**). J. Acoust. Soc. Am. **128**(4), 2112–2126], NSE quantifies information as the change in 1-ms neural firing patterns across frequency. To evaluate the model, data from a study that examined nonlinear frequency compression (NFC) in listeners with SNHL were used because NFC can recode the same input information in multiple ways in the output, resulting in different outcomes for different speech classes. Overall, predictions were more accurate for NSE than CSE. The NSE model accurately described the observed degradation in recognition, and lack thereof, for consonants in a vowel-consonant-vowel context that had been processed in different ways by NFC. While NSE accurately predicted recognition of vowel stimuli processed with NFC, it underestimated them relative to a low-pass control condition without NFC. In addition, without modifications, it could not predict the observed improvement in recognition for word final /s/ and /z/. Findings suggest that model modifications that include information from slower modulations might improve predictions across a wider variety of conditions.
 [http://dx.doi.org/10.1121/1.4934731]

[FJG] Pages: 3061–3072

I. INTRODUCTION

A variety of signal processing algorithms, including speech enhancement and frequency lowering, have been developed for individuals with sensorineural hearing loss (SNHL) in an attempt to partially restore degraded or absent speech cues (e.g., Alexander *et al.*, 2011; Alexander, 2013; Koning and Wouters, 2012). Because signal processing techniques like these use more than simple gain and attenuation to recode speech, traditional or modified audibility-based models of speech intelligibility [e.g., Speech Intelligibility Index (SII); ANSI, 2007] cannot fully capture the change in potential information associated with the increased saliency or distortion of particular speech cues. The purpose of this research is to develop a neural analog of the Cochlear-Scaled Entropy (CSE) model (Stilp *et al.*, 2010), developed in part by J. M. Alexander, which is designed to quantify information in the acoustic signal in terms of how the excitation pattern along the cochlea changes over time in response to a stream of speech. This property makes a model like CSE a possible alternative to traditional models when describing the information available in a speech-enhanced signal where dynamic features have been exaggerated/diminished or where speech energy has been translocated into frequency regions that had none before (i.e., "frequency lowering"). CSE uses concepts from Shannon information theory (Shannon, 1948), which states that information transfer occurs when uncertainty is reduced (in other words, "when the unknown becomes known"). When formally quantified in terms of bits, the measure of uncertainty is "entropy." Speech is a signal characterized by significant acoustic redundancy across time and frequency, however, degradation of the signal during transmission and/or sensory transduction reduces redundancy and threatens robust communication by reducing or eliminating the perceptual saliency of individual speech cues. Where signal redundancy is low, it is less predictable and uncertainty/entropy is high.

Consequently, entropy can be a direct measure of the *potential* speech information that can be conveyed through a communication medium or device.

In order for entropy to be a useful measure of speech information, a conceptual framework of the phenomenological processes involved needs to be applied. This conceptual framework can take many forms depending on the level of processing one chooses to explore. For speech, the processing may be the extraction of linguistic units such as phonemes (Stilp, 2011), syllables, or whole words (Shannon, 1951) at the cortical level. At the level of the auditory periphery, the processing may be the displacement of the basilar membrane (e.g., Stilp *et al.*, 2010) or the pattern of neural firing in the auditory nerve (AN) or anywhere along the auditory pathway. Therefore, decisions need to be made about how one chooses to represent the processing at each level. These decisions have implications for what exactly is defined as uncertain, hence how "information" is manifested (that is, how bits are counted). With respect to bit representations at the auditory

a)Electronic mail: alexan14@purdue.edu

periphery, the definition of signal uncertainty needs to respect processes involved in sensory transduction, with auditory filter tuning being the minimum.

The first step in the computation of CSE is to create a time-frequency representation of the speech (Stilp and Kluender, 2010; Stilp *et al.*, 2010). Time is divided into non-overlapping segments and the short-term spectrum is computed. Individual spectral components are then grouped into non-overlapping frequency bins corresponding to a single ERB (equivalent rectangular bandwidth; Glasberg and Moore, 1990), thereby generating "spectral slices." Implicit with quantifying information at the level of the auditory periphery is the premise that sensory systems are efficient information processors that respond best to changes in the signal across time and frequency (Kluender *et al.*, 2003). Therefore, the next step is to characterize the degree to which each spectral slice is dissimilar from one or more preceding slices (Stilp and Kluender, 2010). Dissimilarity is estimated by the Euclidean distance between successive spectral slices, or across a moving average of spectral slices, and is a proxy for the information-theoretic measure of entropy.

CSE models are accurate at predicting speech across a variety of conditions. For example, CSE has been shown to have a close association with sentence intelligibility under a variety of temporal distortions that an audibility-based model would be insensitive to, including conditions that could not be explained by a temporal envelope model based on the modulation transfer function (Stilp *et al.*, 2010). Furthermore, Stilp and Kluender (2010) showed that speech intelligibility was significantly degraded when groups of spectral slices that were identified as "high CSE" were replaced by speech-shaped noise, whereas intelligibility was largely maintained when an equal number of low-CSE spectral slices were replaced. These findings were replicated using noise-vocoded speech (CSE_{CI}) in an attempt to simulate the speech signal after processing with a cochlear implant, CI (Stilp *et al.*, 2013; Stilp, 2014). Intelligibility was poorer when high-CSE_{CI} parts of sentences were replaced with noise versus when low-CSE_{CI} parts were replaced with noise. Overall, the findings from these studies support the idea that information-bearing change plays a role in speech perception.

As the name suggests, CSE is based on the cochlear level of auditory processing. However, there is evidence that neural processes further up the auditory pathway successively alter the input signal by selectively responding to elements high in information (e.g., Chechik *et al.*, 2002). Adaptation is one such phenomenon present at the level of the AN. Adaptation is the decrease in the neural firing rate in response to constant stimulation. That is, firing rate decreases as entropy decreases since an unchanging stimulus rapidly ceases to be novel. Studies have suggested that perception of speech is enhanced because neurons adapt quickly after initial stimulation, thereby making un-adapted neurons with different characteristic frequencies (CFs) relatively more responsive to onsets (Delgutte, 2002). This collectively results in a pattern of AN discharge peaks that correspond to spectro-temporal regions rich in phonetic content. In other words, adaptation may improve the neural representation of spectral change (entropy) between sequential speech segments (Kluender *et al.*, 2003).

Therefore, the goal of this study is to add to the CSE model by including a level of neural processing. By using the Zilany *et al.* (2014) phenomenological model of the human AN to describe the time-frequency representation of speech at the level of the inner hair cell (IHC)-AN synapse, the new entropy measure, called Neural-Scaled Entropy (NSE), is sensitive to changes in the representation of speech at the neural level. NSE is designed to capture the effects of adaptation, suppression, and other nonlinearities, thereby allowing one to model the effects of SNHL beyond audibility and simple filter broadening.

Perceptual data from Alexander (2012) were used to evaluate the association between NSE and speech intelligibility in conditions where spectral features had been deliberately increased and decreased in perceptual saliency using a common form of frequency lowering in hearing aids. In general, frequency-lowering algorithms in hearing aids are intended for individuals with hearing impairment who have difficulty perceiving high-frequency speech information with conventional amplification. These algorithms use all or part of the aidable low-frequency spectrum to code part of the inaudible high-frequency spectrum. Often, this entails "moving" part of the low-frequency spectrum that would otherwise be aided normally to make room for the additional high-frequency information (i.e., compression techniques) or entails superimposing the recoded high-frequency information on top of the existing low-frequency spectrum (i.e., transposition techniques). See Simpson (2009) and Alexander (2013) for a review of frequency-lowering techniques and associated research findings. With nonlinear frequency compression (NFC), the input spectrum is divided into two bands at a nominal frequency designated as the "start frequency," SF. The band below the SF is unaltered whereas the band above the SF is shifted down in frequency. Spectral components closest to the SF undergo the least frequency shifting on a linear scale and those furthest from the SF undergo the greatest frequency shifting (Simpson *et al.*, 2005). The exact amount of frequency shifting and the subsequent reduction in bandwidth (BW) above the SF depends on the nominal "compression ratio" (CR) with higher ratios corresponding to greater frequency shifts. The particular implementation of NFC simulated by Alexander (2012) lowered frequency on a log scale (Simpson *et al.*, 2005), with the nominal CR being about equal to the reduction in spectral resolution in terms of auditory filter BW (Alexander, 2013). For example, a 2:1 CR meant that information that normally spanned two auditory filters in the un-impaired ear before processing only spanned one auditory filter after processing.

Experiments on NFC are suitable for evaluating NSE because the increase in information associated with re-introducing high-frequency speech cues into the audible frequency range also comes with increased risk of distorting information from low-frequency speech cues including suprasegmental and indexical cues (Alexander, 2013). It follows that the integrity of information after recoding with NFC is not only influenced by the SF and the input BW, but

also on the distribution of low- and high-frequency information in the individual speech sounds being processed. For example, Alexander (2012) found that a low SF (1.6 kHz) was detrimental to recognition of vowels, which relies heavily on formant frequencies (e.g., Ladefoged and Broadbent, 1957), and recognition of word medial consonants which are cued to varying extents by formant transitions of the surrounding vowels (e.g., Sussman *et al.*, 1991; Dorman and Loizou, 1996). It was also found that the effect of CR was greatest at the lowest SF because higher CRs often resulted in progressively lower recognition scores for these phonemes, whereas there was very little effect of changing CR at higher SFs. In other words, where frequency compression occurred seemed to be more important than how much of it there was. On the other hand, for some stimulus tokens, especially /s/ and /z/ with an initial /i/ spoken by female talkers, recognition was best with the 1.6 kHz SF and a high CR, which allowed for a greater input BW to be made available to the listeners after lowering. This was likely because frication information was spectrally diffuse and was therefore less dependent on precise frequency content. In contrast, alteration of the primary formants, which originated at low to mid frequencies, was not as perceptually resilient.

These findings have important implications for how the proposed model, an extension of CSE, is predicted to behave with speech processed by NFC. That is, because CSE utilizes a quasi-logarithmic frequency scale it gives more weight to lower frequencies, which correspond to the formant regions. As a natural formant peak changes frequency across time, it transverses several auditory filters, resulting in high entropy. If the extent of this change is reduced by NFC, it should have a greater effect on entropy compared to, for example, a reduction in the BW of frication, which tends to naturally fall in relatively broader high-frequency filters. Therefore, it is hypothesized that the entropy measures (CSE and NSE) will predict maximum speech degradation for the most aggressive NFC settings (low SF and high CR), and best speech recognition for the least aggressive NFC settings (high SF and low CR) for phonemes that rely to a significant extent on low-frequency formants and formant transitions and will predict the opposite pattern for phonemes that rely to a significant extent on high-frequency frication.

II. METHODS

A. Perceptual data

Model data were compared to perceptual data from Alexander (2012). Fourteen listeners (6 male, 8 female) aged 47–83 yrs (median = 70 yrs), with mild to moderate SNHL (Fig. 1) identified stimuli that were low-pass filtered at 3.3 kHz to simulate a severe-to-profound restriction in audible BW. Stimuli consisted of three types of nonsense syllables that were mixed with speech-shaped noise. First, consonant recognition was tested using 240 vowel-consonant-vowel (VCV) syllables presented in speech-shaped noise at 10 dB signal-to-noise ratio (SNR). These included 20 consonants (/p, t, k, b, d, g, f, θ, s, ʃ, v, z, dʒ, tʃ, m, n, l, r, w, y/) in 3 vowel contexts (/ɑ, i, u/) spoken by 2 male and 2 female adult talkers. Second, vowel recognition was tested

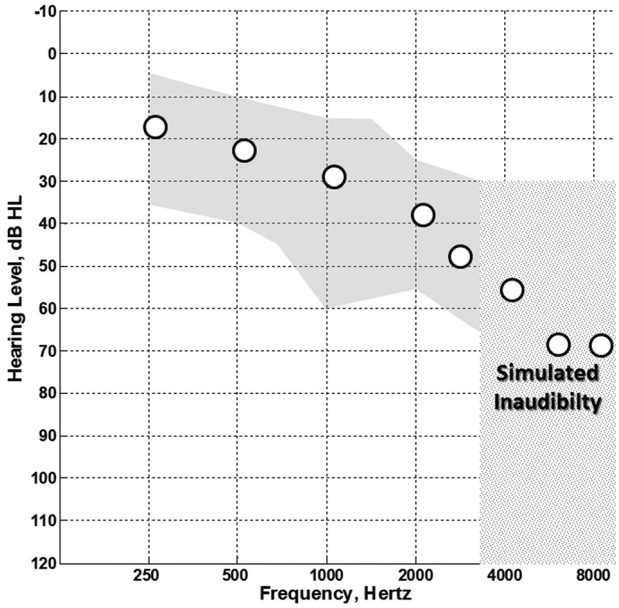

FIG. 1. Mean audiometric thresholds for the listeners in the perceptual experiment (Alexander, 2012) used to evaluate the NSE model. The gray-shaded area represents the range of individual thresholds and the stippled area represents the frequency range above 3.3 kHz where stimuli were low pass filtered to create a moderately-severe to profound BW restriction.

using 144 /h/-vowel-/d/ (hVd) syllables from the Hillenbrand *et al.* (1995) database presented in speech-shaped noise at 5 dB SNR. These included 12 vowels (/i, ɪ, e, ɛ, æ, ɑ, ɔ, o, ʊ, u, ʌ, ɝ/) spoken by 4 adult males, 4 adult females, 2 boys, and 2 girls. Third, recognition of speech stimuli with high-frequency content was tested using 108 fricatives and affricates (/iC/) presented in speech-shaped noise at 10 dB SNR. These included four spoken renditions of seven fricatives (/s, z, ʃ, f, v, θ, ð/) and two affricates (/tʃ, dʒ/) by three female talkers (cf. Stelmachowicz *et al.*, 2001; Alexander *et al.*, 2014).

In a 2×3 design, two NFC SFs, about 1.6 and 2.2 kHz, were crossed with 3 input BWs, about 5, 7, and 9 kHz (Fig. 2). That is, the speech spectrum ranging from the SF to the upper frequency in the input BW was nonlinearly compressed so that the latter was lowered to 3.3 kHz, the maximum amplified frequency. For a fixed SF, increases in audible input BW were accomplished by increases in the CR. NFC was carried out before amplification using an algorithm based on Simpson *et al.* (2005). Briefly, overlapping 128-point Fast Fourier Transforms (FFTs) were computed every 32 samples (1.45 ms) and used to estimate instantaneous frequency. Instantaneous frequencies encompassing the input band above the SF were synthesized at lower frequencies using phase vocoding; with frequency-reassignment determined using Eq. (1) from Simpson *et al.* (2005):

$$F_{\text{out}} = [\text{SF}^{(1-1/\text{CR})}] \times [F_{\text{in}}{}^{(1/\text{CR})}], \quad (1)$$

where F_{out} is the output frequency in kHz, SF is the start frequency kHz, CR is the frequency compression ratio, and F_{in} is the instantaneous input frequency kHz.

Stimuli were presented to listeners via a LynxTWO™ sound card (Lynx Studio Technology, Inc., Newport Beach,

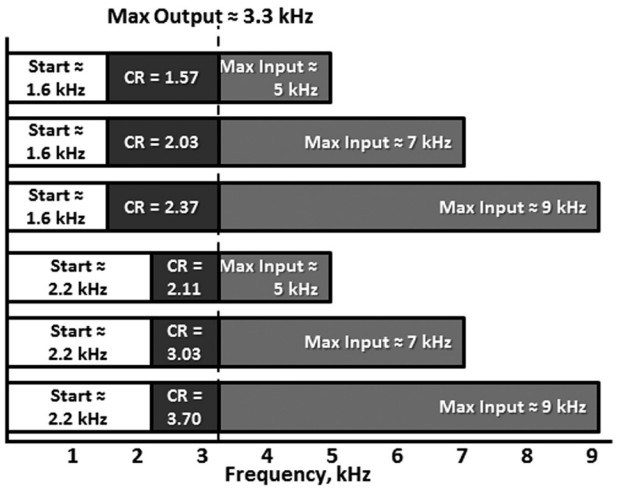

FIG. 2. The six experimental conditions with NFC in which two SFs were crossed with three input BWs. The two gray-shaded areas together represent the range of input frequencies that were subjected to lowering and the dark-shaded area alone represents the range of output frequencies with altered spectral content. All conditions were lowered to a maximum output of 3.3 kHz. For a fixed SF, increases in audible input BW (the highest input frequency, which also corresponded to 3.3 kHz in the output) were accomplished by increases in the CR.

CA) and circumaural BeyerDynamic DT150 headphones (Heilbronn, Germany). Amplification was customized for each listener so that stimuli were presented at output levels prescribed by the Desired Sensation Level algorithm v5.0a (Scollie *et al.*, 2005) using fast-acting, 8-channel wide dynamic range compression in MATLAB (see also Alexander and Masterson, 2015; McCreery *et al.*, 2013, 2014; Brennan *et al.*, 2014, 2015). All stimuli were low-pass filtered at about 3.3 kHz using a 1024-tap finite impulse response filter to simulate a severe high-frequency hearing loss. Stimuli in the control condition were amplified and low-pass filtered, but were not processed with NFC. Because of computational demands in the present study, stimuli that were used as input to the entropy models were amplified using the prescriptive parameters for the average hearing loss of the listeners in the Alexander (2012) study instead of each loss individually.

B. CSE

For comparison with NSE model predictions, CSE was computed using similar methods as Stilp and Kluender (2010). Briefly, each speech stimulus was divided into 16-ms frames. Sixty-six point FFTs were calculated and the output was weighted using ROEX filters (Patterson *et al.*, 1982) to mimic the nonlinear frequency distribution and weighting on the cochlea. As mentioned earlier, this frequency distribution was quasi-logarithmic and corresponded to ERB spacing on the cochlea (Moore *et al.*, 1999). Each band corresponded to one ERB filter up to ERB #26, with center frequencies ranging from 0.03 to 3.54 kHz to include the full BW of the stimuli used in the experiment. To simulate the effects of SNHL on auditory filter BWs, the high-frequency tail of the ROEX filters was gradually increased above 1.42 kHz from normal to 3.65 times normal at 3.54 kHz, with the relative tail width set equal to $2\log_2 F$, where F was the center frequency of the ROEX filter. The low-frequency tail was set to 2 times the width of the high-frequency tail to mimic the upward spread of masking observed in psychophysical estimates of auditory filter shape (e.g., Tyler *et al.*, 1984). Euclidean distances between adjacent 16-ms frames were calculated and averaged across time to compute CSE. To be compatible with the computation of NSE, the initial 50-ms ramp of the stimuli was eliminated during analysis to give the estimate of entropy time to settle from the onset response to the stimulus.

C. Neural time-frequency representation of speech

The Zilany *et al.* (2014) AN model consists of several modules representing the peripheral auditory structures from the middle ear to the AN. Studies have shown that this and previous versions of the model can be used to describe AN responses for a wide variety of stimuli spanning the dynamic range of hearing for both normal and impaired ears (Bandopadhyay and Young, 2004; Tan and Carney, 2006; Hines and Harte, 2010). An advantage for the present purposes is that the AN model includes various aspects of peripheral auditory impairment, including IHC and outer hair cell (OHC) dysfunction, and consequences for perception in terms of loss of audibility, loudness growth, and broadened tuning.

Before processing with the AN model, audio files were passed through a transfer function for a Beyer Dynamic DT150 headphone in order to make the input to the AN model comparable to the signals heard by participants in Alexander (2012). The filtered audio files were processed by the AN model and output at the IHC-AN synapse in spikes/s was used to generate the neural response for the specified CFs at a 100-kHz sampling rate. CFs were sampled at equidistant spaces along the cochlear partition using the Greenwood (1990) function

$$\mathrm{F} = \mathrm{A}(10^{\mathrm{ax}} - \mathrm{k}), \tag{2}$$

where F = frequency in Hz, A = scaling constant between CF and upper frequency limit (165.4), a = slope of the straight-line portion of the frequency-position curve (0.06), k = constant of integration, determined by the lowest audible frequency (0.88), and x = distance in mm. The frequencies of the fibers analyzed ranged from 0.13 to 3.31 kHz, which corresponded to about the lowest allowable frequency in the AN model to about the low-pass filter cutoff frequency of the stimuli used in the experiment. The exact range corresponded to 3.75–22.0 mm on the Greenwood function with a frequency resolution of 0.25 mm, yielding 74 total fibers.

The Zilany *et al.* (2014) model was also used to model the functional consequences of SNHL on AN firing patterns. Scaling factors for the parameters that control the severity of IHC and OHC loss (C_{IHC} and C_{OHC}, respectively) was determined by the model, which used the mean audiometric thresholds of the listeners as input (Fig. 1). Across frequency, the scaling factors for C_{IHC} and C_{OHC} ranged from 0.1 to 0.45, and 0 to 0.30, respectively, with a value of 1.0 representing no loss and 0 representing complete loss. The model was set to simulate the frequency distribution and weighting on the human cochlea. Presentation levels ranged

from 69 to 91 dB sound pressure level and corresponded to the levels that would have been presented to a hypothetical listener who had the same audiometric thresholds as the group mean. Fiber spontaneous rate (SR) was set to simulate low SR fibers. Compared to high SR fibers which saturate at low to moderate input levels, low SR fibers are superior at encoding stimuli at high input levels. They also have an increased resistance to noise masking, possibly due to their narrower BWs and higher thresholds (Silkes and Geisler, 1991). Finally, when modeling SNHL, we expect the firing patterns from the low SR fibers to be more representative of the information carried by the impaired cochlea because SNHL has been shown to result in a disproportionate reduction in high SR fibers (Liberman and Dodds, 1984).

The final step to generating the neural time-frequency representation, the neurogram, involved averaging the simulated neural response over 1-ms time frames. As will be discussed, other time frames were explored, but none yielded nearly as good associations with the data when considered as a whole. The initial 50-ms ramp of each stimulus was eliminated from analysis in order to avoid stimulus onset effects.

D. NSE

To compare how the cochleotopic pattern of neural firing changes across time frames, the Kullback-Liebler Divergence (KLD) can be computed, which quantifies how much one probability distribution differs from another (Johnson *et al.*, 2001; Bandopadhyay and Young, 2004). Using the distribution of spike rates across CFs for one time frame as a prior, KLD computes how many bits are needed to code the neural firing pattern at another time frame. In other words, KLD quantifies how much information remains once mutual information between time frames has been accounted for. Implicit with the computation of KLD is the conversion of spike rate to a unit-less relative probability distribution that describes how much potential information there is in the spike pattern for a given time frame. The problem with this step in the computation is that a silent interval (e.g., a stop closure) will generate a similar probability distribution (equi-probable with maximum entropy) as a high-level broadband noise. In other words, potentially useful information coded by level differences associated with the onsets and offsets of speech, etc., are lost in this process. Therefore, rather than computing the difference in bits between two successive time frames as with a true KLD measure, NSE was computed by calculating the number of bits needed to code the absolute difference in the raw distribution of spike rates across fibers

$$\Sigma P(d_i)\log_2 P(d_i), \text{ where } d_i = |sr_{ij} - sr_{ij-1}|, \quad (3)$$

where P is the probability distribution and d_i is the absolute difference in spike rates (sr) across fibers (i) between frames (j). Essentially, NSE is a measure of dissimilarity (bits of entropy) that describes the degree to which neural responses changed from one time frame to the next. The average NSE across the duration of each stimulus was computed by taking the mean for all segments.

Neural spikes represent a series of cochleotropic distributions. Higher auditory centers transfer information by learning the source of each spike received. An efficient coding strategy is to assume stasis and code only differences. That is, if the neural firing pattern changes very little across time, earlier time frames will closely predict later ones and entropy will be very low. In addition, changes across many fibers, like a lower-frequency formant transition, will have relatively greater entropy than random changes across fewer fibers, like frication, because the distribution of firing across CF changes very little with higher-frequency noise.

III. ANALYSIS

For the purposes of evaluating the NSE model, analyses were carried out for the consonant stimuli according to their manner of articulation, and for the vowels according to tongue advancement. The entropy models were evaluated this way in an attempt to relate differences in observed recognition and differences in model predictions across NFC conditions to the speech acoustics that vary along with these features. That is, in a VCV context, fricatives generally have more high-frequency information and require a greater BW for maximum recognition than stops, followed by liquids and glides (approximants), while affricates and nasals in this context require very little high-frequency information to achieve maximum recognition (Alexander, 2010). It is also well known that the frequency of F_2 varies systematically with the advancement of the tongue during vowel production, with the lowest F_2 frequencies for the back vowels and the highest F_2 frequencies for the front vowels.

Before averaging under each stimulus category, individual percent correct scores for the 14 listeners were converted to rationalized arcsine units (RAUs; Studebaker, 1985) in order to linearize the variance associated with scores reported as proportion correct. Linear regression was carried out against the mean RAU scores for each condition and the corresponding NSE. The slope and intercept obtained from this analysis were used to derive predictions for NSE. R^2 was used to quantify the degree of association between NSE and speech recognition and the Benjamini-Hochberg (1995) False Discovery Rate (FDR) correction was applied to the p-values to control the familywise error rate. Because NSE is built on the framework of CSE, speech recognition was also compared with CSE in the same manner as NSE.

IV. RESULTS

A. Consonant stimuli

Table I shows the R^2 for NSE and CSE across manner of articulation. The 95% confidence intervals for the R^2 from the NSE model are displayed in the bottom row. The features analyzed included stops, affricates, fricatives, liquids and glides (approximants), and nasals. As reported by Alexander (2012), due to their low-frequency emphasis, recognition of the affricates and nasals was relatively homogeneous across conditions with little to no significant paired comparisons. Consequently there was essentially no variance in the data to describe, and R^2 for NSE and CSE for these sound classes

TABLE I. Displayed are the R^2 values for the consonants in VCV nonsense syllables. The three column sets show R^2 across each manner of articulation (stops, fricatives, and liquids & glides, respectively). Within each manner, the first column shows R^2 for comparisons between NSE and the perceptual data while the second column shows comparisons between CSE and the perceptual data. Asterisks denote the level of statistical significance after correcting for multiple comparisons using the FDR (*significant at $p < 0.05$, **significant at $p < 0.01$, ***significant at $p < 0.001$). The last row shows the 95% confidence intervals for the R^2 of the NSE.

	Stops		Fricatives		Liquids and Glides	
	NSE	CSE	NSE	CSE	NSE	CSE
R^2	0.94***	0.55	0.93***	0.39	0.89**	0.50
Confidence intervals	(0.88–1.01)		(0.84–1.01)		(0.77–1.01)	

were unsurprisingly insignificant. Therefore, they will not be discussed further. As indicated by the values for R^2 in Table I, changes in NSE across conditions were closely associated with the observed differences in recognition for stops, fricatives, and liquids and glides, and accounted for around 90% or more of the variance in the data. On the other hand, R^2 for CSE was not significant for any manner of articulation and was outside the confidence intervals for NSE, indicating that NSE was a better predictor of recognition than CSE for each manner of articulation.

To better understand how the predictions of the entropy models compared with consonant recognition and with each other, Fig. 3 uses bar graphs to plot the observed data for stops, fricatives, and liquids and glides [panels (a)–(c), respectively]. Predicted recognition scores from the regressions on NSE and CSE are plotted as light gray squares and dark gray triangles, respectively. Error bars correspond to the standard error of the mean of the observed data and asterisks above each BW correspond to the level statistical significance between the SF pair after applying the FDR correction for all possible paired comparisons. As indicated by Fig. 3, both NSE and CSE accurately predicted the best recognition for the low pass control (LPC) condition, which is plotted as the unfilled bar in each panel. However, the two models differed in how accurately they described the effect of SF on recognition. To help visualize the predicted effects of SF at each input BW, predictions for NSE and CSE are connected by solid and dotted lines, respectively. Whereas NSE accurately predicted improved recognition with the higher SF (light gray bars) compared with the lower SF (dark gray bars) for the different manners of articulation, CSE tended to predict equivalent recognition at the two SFs for the stops and liquids/glides and to predict slightly degraded recognition with the higher SF for the fricatives.

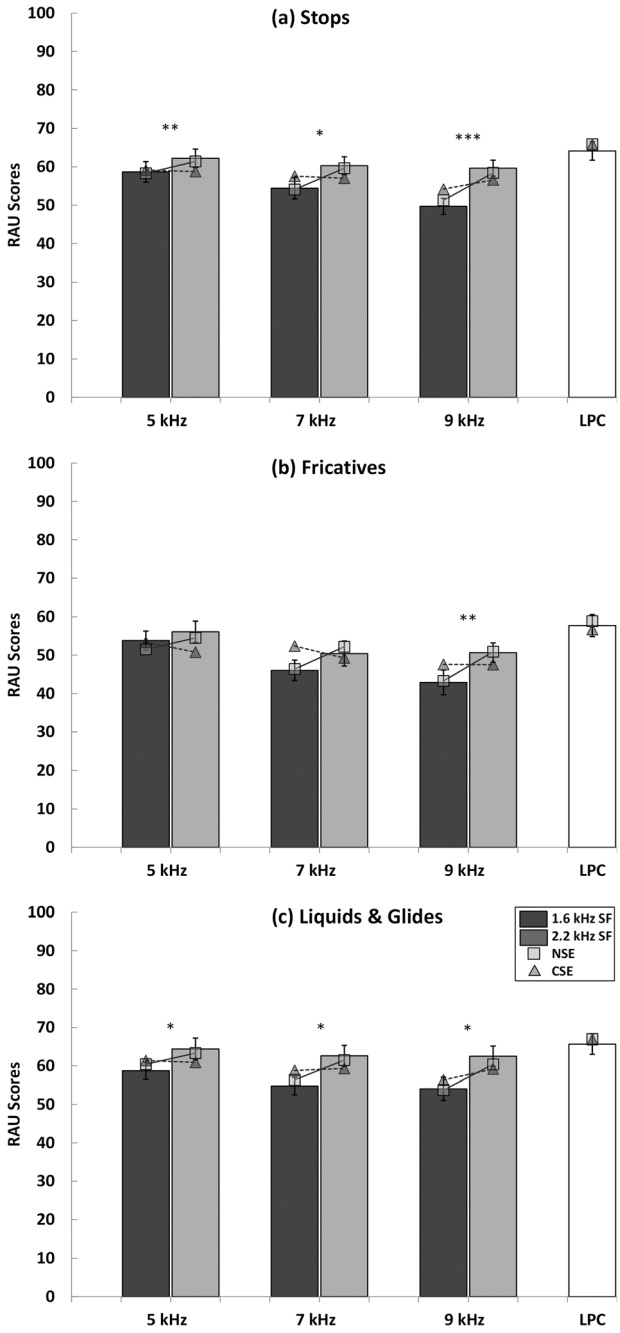

FIG. 3. (a)–(c) show the observed and predicted RAU scores as a function of maximum input BW for the consonant stimuli analyzed by manner of articulation. The dark gray bars represent the low SF (1.6 kHz) conditions and the light gray bars represent the high SF (2.2 kHz) conditions. The last unfilled bar depicts the LPC condition. Corresponding predictions from the NSE and CSE models are indicated by the light gray squares and dark gray triangles, respectively. For a given input BW, solid lines connect the NSE predictions and dotted lines connect the CSE predictions. Error bars correspond to the standard error of the mean of the data and asterisks above each pair correspond to the level statistical significance reported by Alexander (2012), which applied the FDR correction assuming all possible paired comparisons (*significant at $p \leq 0.05$, **significant at $p \leq 0.01$, ***significant at $p \leq 0.001$). (a) Stops; (b) fricatives; (c) liquids and glides.

B. Vowel stimuli

Figure 4 plots the observed data and model predictions for the front, central, and back vowels [panels (a)–(c), respectively]. Predictions from NSE and CSE (squares and triangles, respectively) were similar to each other as indicated by the overlapping symbols in each plot. There were greater differences in the observed recognition scores between conditions for the entropy models to predict with vowels than with consonants, especially as a function of SF for which differences increased as input BW increased. As indicted in the figures, while NSE and CSE predicted the highest recognition for the LPC, they also consistently overestimated it. In addition, the two models generally predicted

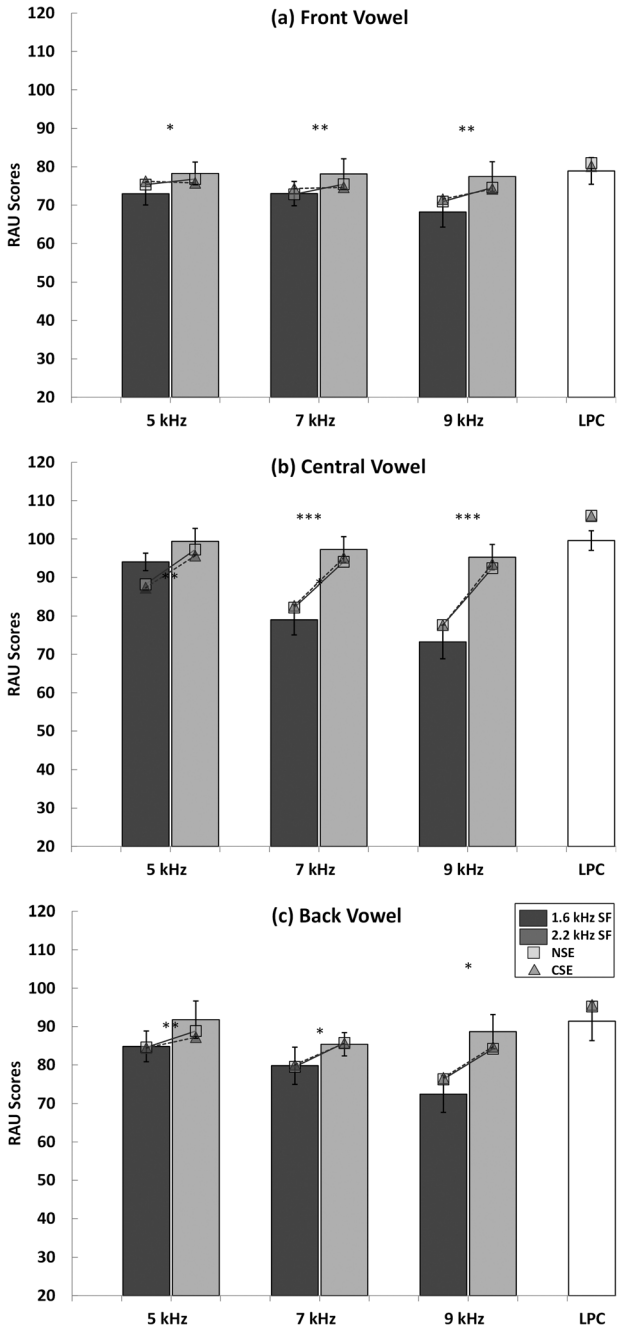

FIG. 4. Observed and predicted RAU scores for the vowel stimuli as a function of tongue advancement are plotted in the same way as Fig. 3: (a) Front vowels; (b) central vowels; (c) back vowels. The ordinate ranges from 20 to 120 RAUs so that the displayed scale covers the same range as for the consonant stimuli.

lower-than-observed recognition for the higher SF, which when combined with the overestimate from the LPC condition suggests a predicted detriment in recognition for these NFC conditions that did not exist before.

The accuracy of the model predictions shown for the vowel data in Fig. 4 are summarized by the regression results provided in Table II. There were significant relationships between NSE and recognition for vowels in each set and significant relationships between CSE and recognition for the back and central vowels. R^2 for CSE failed to reach statistical significance for the front vowels (highest F_2 frequencies),

TABLE II. Displayed are the R^2 values for vowels analyzed across the feature of tongue advancement (front, central, and back), with asterisks denoting statistical significance ($p < 0.05$). The second row shows the 95% confidence intervals for the R^2 of the NSE.

	Front		Central		Back	
	NSE	CSE	NSE	CSE	NSE	CSE
R^2	0.63*	0.43	0.81*	0.79*	0.80*	0.75*
Confidence intervals	(0.29–0.98)		(0.61–1.01)		(0.58–1.01)	

in part because it inaccurately predicted almost equivalent recognition across five of the six conditions with NFC. For the front vowels, CSE only accurately predicted that the highest recognition would be achieved with the LPC and that the lowest would be achieved with the NFC condition with the low SF and the greatest input BW (i.e., the condition with the greatest frequency shifting). While significant, R^2 for NSE was also the lowest for front vowels; therefore, the confidence intervals for NSE included the R^2 values for CSE for all of the vowel sets, which suggests that the accuracies of the models were statistically equivalent.

C. High-frequency stimuli

The high-frequency stimuli consisted of seven fricatives and two affricates spoken by three female talkers in the /iC/ context, and of these, Alexander (2012) reported that only /s/ and /z/ showed significant effects across conditions. The observed effects for both /s/ and /z/ were somewhat opposite of the effects for the consonant and vowel stimuli because recognition for the LPC condition was much lower relative to the best conditions with NFC, which happened to be the conditions that were most detrimental to recognition of the VCVs and hVds. Because the observed data for /s/ and /z/ were divergent from what was observed and predicted for the VCV and hVd stimuli, it provided a good test case to see if the entropy models would also predict better recognition with the NFC conditions compared with the LPC condition. The pattern of NSE across conditions was similar to that reported for the VCVs and hVds. The LPC condition had the highest NSE, and the 1.6-kHz SF conditions with 7- and 9-kHz BWs had the lowest NSE, which was opposite of the observed data. This resulted in significant negative correlations for /s/ [$r = -0.91$, $R^2 = 0.83$, $p < 0.01$] and for /z/ [$r = -0.88$, $R^2 = 0.78$, $p < 0.01$].

As indicated in Fig. 5, while the CSE prediction for the LPC condition was substantially overestimated for both phonemes, it did predict better recognition for some of the NFC conditions, which did not occur with the original NSE predictions. This observation prompted an exploration of the parameters that generated the NSE values, which suggested that the duration of the analysis time frame was a critical factor. In particular, relatively favorable predictions were obtained when NSE was evaluated using a moving average ($\tau = 5$) across 16-ms frames as proposed by Stilp and Kluender (2010) for CSE when analyzing entropy in longer duration sentences. As shown in Fig. 5, by changing the NSE model, the relative prediction for the LPC condition

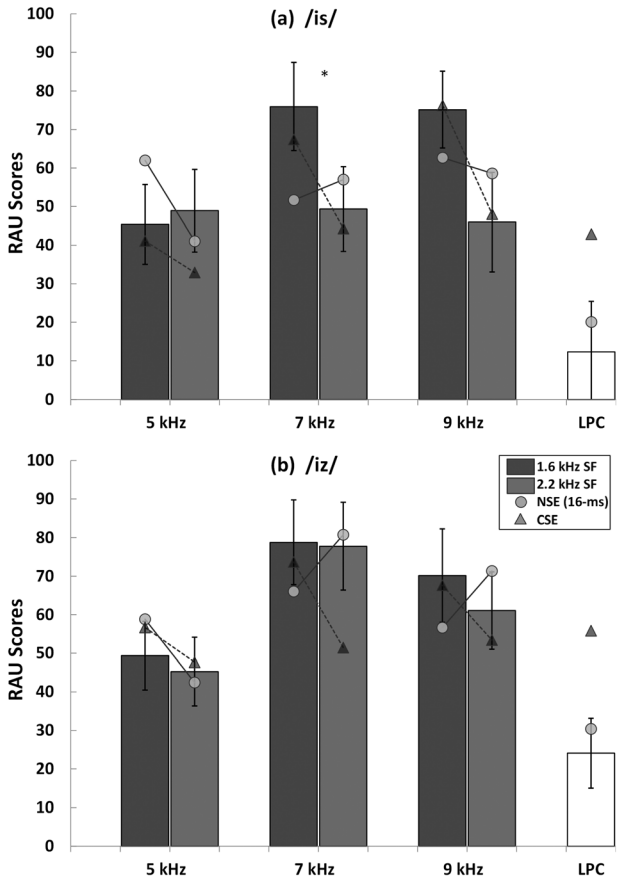

FIG. 5. Observed and predicted RAU scores for (a) /is/ and (b) /iz/ are plotted in the same manner as Fig. 4, except that gray-filled circles are used to plot NSE predictions since they were computed differently than the other NSE predictions and were based a moving average across 16-ms time frames.

went from being greater than (as with the original NSE predictions, not shown) to being less than the predictions for all of the NFC conditions, just as was true for the observed data. Table III shows that R^2 for the modified NSE model was marginally significant for /s/ ($p = 0.07$) and statistically significant for /z/.

While the NSE predictions came close to capturing the magnitude of the differences between the NFC conditions and the LPC condition, CSE did better at describing the variation in recognition scores across conditions than NSE for /s/, but not for /z/. The interpretation of these results is not entirely clear because there was substantial individual variability in the observed data, as indicated by the large standard error bars, so that many of the paired comparisons between NFC conditions were not statistically significant.

TABLE III. Displayed are the R^2 values for the high-frequency tokens /is/ and /iz/ for the NSE model when computed using a moving average across 16-ms time frames and for the CSE model when computed using previous methods. The asterisk denotes statistical significance ($p < 0.05$). The second row shows the 95% confidence intervals for the R^2 of the NSE.

	/is/		/iz/	
	NSE	CSE	NSE	CSE
R^2	0.51	0.52	0.75*	0.22
Confidence intervals	(0.1–0.92)		(0.50–1.00)	

V. DISCUSSION

A. NSE model predictions

The NSE model was developed to quantify how speech information is affected by hearing impairment and various signal processing strategies that alter the content of the speech signal beyond simple gain adjustments. Data from a study that examined NFC in a group of listeners with SNHL (Alexander, 2012) were used to test and refine the NSE model. NFC is suited for assessing the performance of the model because there are multiple ways that the same input BW of information can be recoded into a narrower output BW and the different ways have been demonstrated to result in different outcomes, depending on the sounds being lowered. For example, Alexander (2012) reported that sounds characterized by low-frequency formants, especially vowels, were distorted the most by low SFs and higher CRs. On the other hand, spectrally diffuse, high-frequency sounds, such as fricatives and affricates in the word-final position, were found to be more perceptually resilient to low SFs and high CRs.

The initial evaluation of the NSE model as presented in this report indicates that it offers promise for making predictions for recognition of spectrally altered speech. It is important to recognize that evaluation of the model was based on mean data of broad speech classes from a group of listeners with SNHL. Ideally, predictions will be able to be made on an individual token-by-token basis for individual listeners; however, the purpose of averaging was to reduce variability that is typical of human perceptual studies. Given that the majority of individual thresholds through 3 kHz were within 10 dB of the mean audiogram (Fig. 1) used to generate the stimuli and to compute the NSE, it is doubtful that there would be any substantial changes in the NSE and subsequent predictions. The purpose of analyzing the data by speech class was to identify general patterns that would provide valuable information about the sort of conditions for which NFC was beneficial or detrimental. Likewise, it made sense to evaluate the accuracy of the NSE model in the same manner.

Existing models that predict speech recognition for conventional amplification, such as the SII, primarily account for a loss of audibility (hence, information) via threshold shifts and the restoration thereof. They do not account for nonlinear distortions such as filter broadening resulting from SNHL. Therefore, they are likely to be less successful when tested with a wide range of NFC settings as was tested in this study. Proposals have been made to estimate changes in speech recognition with NFC using audibility-based models, such as the modified SII (Bentler *et al.*, 2011; McCreery *et al.*, 2014). The modified SII is based on the assumption that improvements in speech recognition will be proportional to how much of the input BW of speech is made "audible" after frequency lowering. Following this assumption, relative to the LPC condition, the modified SII would predict an improvement in recognition with NFC across all speech stimuli and all combinations of SF and CR. In addition, it would predict that recognition would increase as CR is increased because of its relationship with input BW in this

study and it would predict no effect for SF across conditions with equal input BW.

Contrary to audibility-based models, which predict an increase in speech recognition for all NFC conditions, distortion-based models of speech perception only predict a decrease in recognition for all speech sound classes. An example of a distortion-based model that has been applied to speech processed by a variant of NFC is the Hearing Aid Speech Perception Index (HASPI; Kates and Arehart, 2014). The HASPI compares the envelope and temporal fine structure outputs of an auditory model for a reference signal set to normal hearing to the outputs of the model for the test signal that incorporates hearing loss. As implemented by Kates and Arehart (2014), the HASPI only predicted detriments to speech recognition following processing with NFC because the reference was the wideband signal. However, it is possible that the HASPI will predict improvement for the high-frequency /is/ and /iz/ in Alexander (2012) if the NFC-processed speech is compared to the HASPI for the LPC condition. The reason for this may be related to the reason why modifications to the NSE model had to be made by using longer time frames in order for it to predict improved recognition for these stimuli. While it seems clear that temporal fine structure cues will be adversely affected by all forms of NFC, it may also be the case that when NFC improves perception for certain stimuli, it is due to greater access to the slower modulations of speech, namely the temporal envelope.

Lowered temporal envelope information from frication can inform the hearing aid user about (a) the presence of a high-frequency speech event; (b) its subclass (a fricative) based on its noisy sound quality; (c) the relative duration of the original sound; and (d) the relative intensity and changes therein over the duration of event (i.e., the amplitude

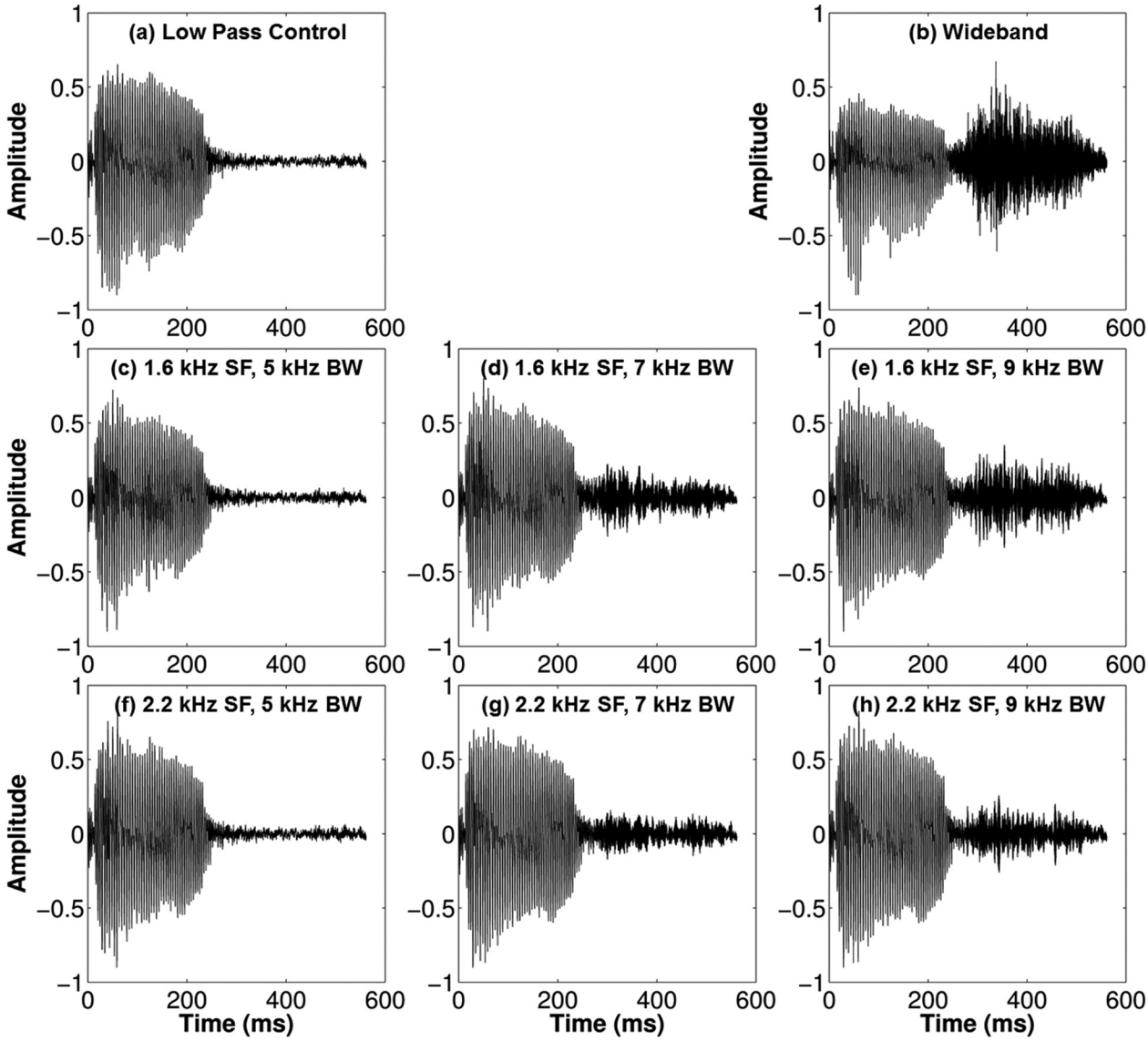

FIG. 6. Resultant time waveforms are plotted for an exemplar of the stimulus /is/ spoken in quiet after processing it in the same way as the experimental conditions. Amplitude is on the ordinate and time in milliseconds is on the abscissa. The LPC condition (a) shows very little energy associated with the frication for /s/ (~250–560 ms) compared to the wideband condition (b). (c)–(e) show the waveforms for NFC conditions with a 1.6-kHz SF and (f)–(h) show the waveforms for NFC conditions with a 2.2-kHz SF. Increasing the input BW of the lowered signal from 5–kHz [(c) and (f)] to 7 kHz [(d) and (g)] to 9 kHz [(e) and (h)] progressively reintroduces the temporal features of the missing phoneme.

envelope). To demonstrate how the availability of temporal envelope information changes across conditions, Fig. 6 displays the time waveforms for the phoneme /is/ in quiet for the LPC condition, the wideband signal, and each of the NFC conditions. It can be seen that there is almost a complete absence of acoustic energy from the frication in the second half of the time waveform in the LPC condition [Fig. 6(a)] compared with wideband signal [Fig. 6(b)]. Furthermore, it can be seen that the time waveforms for the two conditions with 5-kHz input BW [Figs. 6(c) and 6(f)] were similar to the LPC and that the time waveforms for the two conditions with 9-kHz input BW were more similar to the wideband signal especially for the 1.6 kHz SF [Fig. 6(e)] compared with the 2.2 kHz SF [Fig. 6(h)].

From the example shown in Fig. 6, it seems likely that predictions based on temporal envelope alone will be greater for NFC conditions that make more of it available for sounds like /is/ and /iz/ compared with the LPC condition. Because the HASPI considers information from temporal envelope in addition to temporal fine structure, it too might be able to predict the pattern in the observed data. Likewise, it may be for this reason that CSE, which computed entropy from 16-ms time frames, predicted higher recognition for /is/ and /iz/ for some NFC conditions compared with the LPC condition while the original NSE computations, which were based on 1-ms time frames, predicted the opposite pattern. That is, in terms of temporal fine structure NSE may have predicted *less* entropy for frequency-lowered frication because the introduction of low-frequency noise created less variation in the distribution of firing across fibers from frame to frame. Therefore, lengthening the NSE time frames to 16-ms and extending the comparisons over time made the model sensitive to the observed changes in temporal envelope and significantly improved predictions.

Compared to consonant recognition, NSE predictions for vowel recognition were less accurate (average $R^2 = 0.92$ for consonants vs average $R^2 = 0.75$ for vowels). Removing the overestimated LPC condition from the vowel analyses improved the predictions (average $R^2 = 0.86$), which suggests that a primary source of error was the relative difference in NSE for the LPC compared with the NFC conditions. Mathematically, at least the same improvements can be achieved by adjusting the NSE for the LPC condition downward or by adding a constant to each NFC condition. It is speculated that instead of the model overestimating recognition for the LPC condition, that it underestimated recognition for the all the NFC conditions. This scenario is plausible if there are other sources of information that can partially offset the degradation incurred at the fine structure level. As just discussed, combining information from other, slower, modulations may be the key to improving the NSE model, although the best way to do this needs to be explored in more detail.

Including one or more additional sources of information in the NSE model may also help address other limitations of the current project that are related to the apparent lack of predicted improvement for NFC-processed speech. One such limitation concerns the possibility that the current predictions may not hold if listeners are given extended experience with each NFC condition because they might learn to make better use of the new and altered cues over time and show improved recognition for certain phonemes (e.g., Wolfe *et al.*, 2010, 2011; Kuk *et al.*, 2009). Having an additional source of information in the model that is more representative of these cues (e.g., temporal envelope) may help to predict a benefit for other phonemes and conditions that were not actualized by the listeners in this experiment. Given that the entropy models are intended to quantify the amount of *potential* information in the speech signal and not necessarily whether listeners actually use this information, it is important that the model, at minimum, be able to document improvements in recognition where they exist.

B. CSE model predictions

While CSE model predictions were similar to NSE model predictions for the hVds, they were significantly poorer for the VCVs. The reason for this may be due to differences in the time scales that each operates and/or differences in the auditory representation of the stimuli. When CSE was computed using normal-shaped auditory filters instead of broadened filters, model predictions were generally worse, which suggests that incorporating this aspect of SNHL was an important feature for both models. Other than the auditory level at which each model operated, there were fundamental differences in the way CSE and NSE were computed that might have led to the differences in their ability to predict the observed data. First, as mentioned, the time and frequency resolution of the auditory representation was different between the two models. Whereas NSE was computed using 74 fibers and 1-ms frames, CSE was computed using 26 ERB filters and 16-ms frames. To examine whether this difference might explain the difference in the predictive power of the two models, CSE was recomputed using 76 ERB filters from ERB #1 to #26, spaced every 1/3 ERB apart in 1.375 ms time frames (the shortest time frame possible for the 198-point FFT that was used to analyze the spectral slices). Another fundamental difference was that while CSE quantified information (dissimilarity) using the Euclidean distances between smoothed spectra, NSE quantified it using the number bits needed to code the absolute differences in the raw distributions of spike rates across fibers. To examine whether differences in the computation of entropy might also explain differences in their predictive power, CSE for both time frames was computed in terms of bits using Eq. (3).

For the consonants and vowels, respectively, Tables IV and V compare the R^2 values for CSE when computed with 1- and 16-ms time frames and when computed in terms of Euclidean distance and in terms of bits. A comparison between row 1 vs row 2 (Euclidean distance) and between row 3 vs row 4 (bits) in each table indicates that increasing the time and frequency resolution of the CSE analysis did not substantially alter the R^2 values (changes in R^2 ranged from -0.12 to $+0.14$) or their statistical significance. In contrast, an interesting pattern emerges when comparing the effect of using bits to compute CSE on the change in R^2. For both time frames of analysis, the change in R^2 when going

TABLE IV. For the CSE model, displayed are the R^2 values for the consonants in VCV nonsense syllables as a function of manner of articulation. The column labeled "Time Frame" corresponds to the duration of the time frames used in the computation of CSE. The column labeled "Method" corresponds to the method used to compute entropy [Euc. Dist. = Euclidean distance; Bits = method in Eq. (3)]. For reference, the first row is the same data reported in Table I. Asterisks denote the level of statistical significance after correcting for multiple comparisons using the FDR (*significant at $p < 0.05$).

Time frame	Method	Stops	Fricatives	Liquids/Glides
16 ms	Euc. Dist.	0.55	0.39	0.50
1 ms	Euc. Dist.	0.53	0.53	0.51
16 ms	Bits	0.69*	0.76*	0.56
1 ms	Bits	0.63	0.75*	0.50

from Euclidean distance to bits is related to the frequency region of primary importance associated with each speech class; from highest to lowest frequency, the average change in R^2 was +0.30 (fricatives), +0.12 (stops), +0.03 (liquids and glides), −0.14 (back vowels), −0.20 (central vowels), −0.22 (back vowels).

In general, the improvement in R^2 for fricatives and stops when computing CSE with bits vs Euclidean distance reflected better predictions for the modest increase in recognition when going from the lower SF to the higher SF (recall that the Euclidean distance metric predicted a decrease or no change). On the other hand, the decrement in R^2 for the vowels reflected smaller predicted differences between the two SFs compared with the large differences in the observed data. The pattern of predicted scores vary between the two methods because the bit computation took the absolute difference between two spectral slices, while Euclidean distance took the square of the differences which had the effect of exaggerating changes in the signal that passed through each filter. For NFC-processed consonants, the change in entropy associated with a decrease in SF represented a tradeoff between making frication more informative (by spreading it out over a greater number of auditory filters) and preserving formant transitions. It is possible that Euclidean distance overestimated recognition for the lower SF by placing too much weight on the spectral changes associated with the frication. For NFC-processed vowels, the information-bearing change was confined to the formants; therefore, Euclidean distance may have been more favorable than bits for capturing the large differences in entropy (hence recognition) between the two SFs.

In summary, the small differences in R^2 when CSE was computed over 1- and 16-ms time frames and the varied change in R^2 when going from computation using Euclidean distance vs bits, lend support to the idea that the differences in the predictive power of the NSE and CSE models were mostly due to unspecified processes that occur at the level of the AN.

TABLE V. For the modified CSE models, displayed are the R^2 values for front, central, and back vowels. See the description of Table IV for more details.

Time frame	Method	Front	Central	Back
16 ms	Euc. Dist.	0.43	0.79*	0.75*
1 ms	Euc. Dist.	0.37	0.70*	0.65*
16 ms	Bits	0.28	0.58	0.54
1 ms	Bits	0.24	0.52	0.42

VI. CONCLUSION

The NSE model best described the observed degradation and the lack thereof, for recognition of consonants in a VCV context that had been processed in different ways by NFC. Identified areas where the model could be improved include recognition of vowels and recognition of consonants loaded with high-frequency content (e.g., final /s/ and /z/ with the leading vowel /i/ as spoken by female talkers). The fact that the model underestimated recognition of NFC-processed speech for both of these stimulus sets suggests that it needs to consider other sources of information that can be used to maintain or enhance perception when information from temporal fine structure is distorted by the signal processing and/or SNHL. It was proposed that including another stream of information that is sensitive to information from slower modulations might improve model predictions across a wider variety of conditions.

Future work will have to consider how to weigh the different sources of information in a unified model and whether certain individual or stimulus factors influence this weighting.

ACKNOWLEDGMENTS

The authors thank Michael Heinz and Keith Kluender for their recommendations during the development of the NSE model and Christian Stilp and three anonymous reviewers for their useful suggestions on an earlier version of this manuscript. This research was funded by NIDCD Grant No. RC1 DC010601.

Alexander, J. M. (**2010**). "Maximizing benefit from nonlinear frequency compression," *4th Phonak Virtual Audiology Conference*.
Alexander, J. M. (**2012**). "Nonlinear frequency compression: Balancing start frequency and compression ratio," *39th Annual Scientific and Technology Conference of the American Auditory Society*, Scottsdale, Arizona.
Alexander, J. M. (**2013**). "Individual variability in recognition of frequency-lowered speech," Semin. Hear. **34**(2), 86–109.
Alexander, J. M., Jenison, R. L., and Kluender, K. R. (**2011**). "Real-time contrast enhancement to improve speech recognition," PLoS One **6**(9), e24630.
Alexander, J. M., Kopun, J. G., and Stelmachowicz, P. G. (**2014**). "Effects of frequency compression and frequency transposition on fricative and affricate perception in listeners with normal hearing and mild to moderate hearing loss," Ear Hear. **35**, 519–532.
Alexander, J. M., and Masterson, K. (**2015**). "Effects of WDRC release time and number of channels on output SNR and speech recognition," Ear Hear. **36**, e35–e49.
ANSI (**2007**). ANSI S3.5-1997 (R2007), *Methods for Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).
Bandopadhyay, S., and Young, E. D. (**2004**). "Discrimination of voiced stop consonants based on auditory nerve discharges," J. Neurosci. **24**(2), 531–541.
Benjamini, Y., and Hochberg, Y. (**1995**). "Controlling the false discovery rate: A practical and powerful approach to multiple testing," J. Roy. Stat. Soc. B **57**, 289–300, available at http://www.jstor.org/stable/2346101.

Bentler, R., Cole, W., and Wu, Y. H. (**2011**). "Deriving an audibility index for frequency-lowered hearing aids," *38th Annual Scientific and Technology Conference of the American Auditory Society*, Scottsdale, Arizona.
Brennan, M. A., McCreery, R., Kopun, J., Alexander, J. M., Lewis, D., and Stelmachowicz, P. G. (**2014**). "Paired comparisons of nonlinear frequency compression, extended bandwidth, and restricted bandwidth hearing-aid processing for children and adults with hearing loss," J. Am. Acad. Audiol. **25**, 983–998.
Brennan, M. A., McCreery, R., Kopun, J., Hoover, B., Alexander, J. M., Lewis, D., and Stelmachowicz, P. G. (**2015**). "Masking release in children with hearing loss when using amplification," J. Am. Acad. Audiol (in press).
Chechik, G., Globerson, A., Tishby, N., Anderson, M. J., Young, E. D., and Nelken, I. (**2002**). "Group redundancy measures reveal redundancy reduction in the auditory pathway," in *Advances in Neural Information Processing Systems*, edited by T. G. Dietterich, S. Becker, and Z. Ghahramani (MIT Press, Cambridge, MA), pp. 173–180.
Delgutte, B. (**2002**). "Auditory neural processing of speech," in *The Handbook of Phonetic Sciences*, edited by W. Hardcastle and J. Laver (Blackwell, Oxford), pp. 507–538.
Dorman, M. F., and Loizou, P. C. (**1996**). "Relative spectral change and formant transitions as cues to labial and alveolar place of articulation," J. Acoust. Soc. Am. **100**, 3825–3830.
Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**(6), 2592–2605.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.
Hines, A., and Harte, N. (**2010**). "Speech intelligibility from image processing," Speech Commun. **52**, 736–752.
Johnson, D. H., Gruner, C. M., Baggerly, K., and Seshagiri, C. (**2001**). "Information—Theoretic analysis of neural coding," J. Comput. Neurosci. **10**, 47–69.
Kates, J., and Arehart, K. (**2014**). "The Hearing-Aid Speech Perception Index (HASPI)," Speech Commun. **65**, 75–93.
Kluender, K. R., Coady, J. A., and Kiefte, M. (**2003**). "Sensitivity to change in perception of speech," Speech Commun. **41**, 59–69.
Koning, R., and Wouters, J. (**2012**). "The potential of onset enhancement for increased speech intelligibility in auditory prostheses," J. Acoust. Soc. Am. **132**, 2569–2581.
Kuk, F., Keenan, D., Korhonen, P., and Lau C. (**2009**). "Efficacy of linear frequency transposition on consonant identification in quiet and in noise," J. Am. Acad. Audiol. **20**(8), 465–479.
Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.
Liberman, M. C., and Dodds, L. W. (**1984**). "Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves," Hear. Res. **16**, 55–74.
McCreery, R. W., Alexander, J. M., Breenan, M. A., Hoover, B., Kopun, J., and Stelmachowicz, P. G. (**2014**). "The influence of audibility on speech recognition with nonlinear frequency compression for children and adults with hearing loss," Ear Hear. **35**(4), 440–447.
McCreery, R. W., Breenan, M. A., Hoover, B., Kopun, J., and Stelmachowicz, P. G. (**2013**). "Maximizing audibility and speech recognition with nonlinear frequency compression by estimating audible bandwidth," Ear Hear. **34**, e24–e27.
Moore, B., Glasberg, B. R., and Vickers, D. A. (**1999**). "Further evaluation of a model of loudness perception applied to cochlear hearing loss," J. Acoust. Soc. Am. **106**, 898–907.
Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788–1803.
Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Laurnagaray, D., Beaulac, S., and Pumford, J. (**2005**). "The desired sensation level multistage input/output algorithm," Trends Amplif. **9**(4), 159–197.
Shannon, C. E. (**1948**). "A mathematical theory of communication," Bell Sys. Tech. J. **27**, 379–423 and 623–656.
Shannon, C. E. (**1951**). "Prediction and entropy of printed English," Bell Labs Tech. J. **30**(1), 50–64.
Silkes, S. M., and Geisler, C. D. (**1991**). "Responses of 'lower-spontaneous-rate' auditory-nerve fibers to speech syllables presented in noise. I: General characteristics," J. Acoust. Soc. Am. **90**, 3122–3139.
Simpson, A. (**2009**). "Frequency-lowering devices for managing high-frequency hearing loss: A review," Trends Amplif. **13**(2), 87–106.
Simpson, A., Hersbach, A. A., and McDermott, H. J. (**2005**). "Improvements in speech perception with an experimental nonlinear frequency compression hearing device," Int. J. Audiol. **44**, 281–292.
Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., and Lewis, D. E. (**2001**). "Effect of stimulus bandwidth on the perception of /s/ in normal and hearing-impaired children and adults," J. Acoust. Soc. Am. **110**, 2183–2190.
Stilp, C. E. (**2011**). "The redundancy of phonemes in sentential context," J. Acoust. Soc. Am. **130**(5), EL323–EL328.
Stilp, C. E. (**2014**). "Information-bearing acoustic change outperforms duration in predicting intelligibility of full-spectrum and noise-vocoded sentences," J. Acoust. Soc. Am. **135**(3), 1518–1529.
Stilp, C. E., Goupell, M. J., and Kluender, K. R. (**2013**). "Speech perception in simulated electric hearing exploits information-bearing acoustic change," J. Acoust. Soc. Am. **133**(2), EL136–EL141.
Stilp, C. S., Kiefte, M., Alexander, J. M., and Kluender, K. R. (**2010**). "Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentences," J. Acoust. Soc. Am. **128**(4), 2112–2126.
Stilp, C. E., and Kluender, K. R. (**2010**). "Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility," Proc. Nat. Acad. Sci. **107**(27), 12387–12392.
Studebaker, G. A. (**1985**). "A rationalized arcsine transform," J. Speech Hear. Res. **28**, 455–462.
Sussman, H., McCaffrey, H. A., and Matthews, S. A. (**1991**). "An investigation of locus equations as a source of relational invariance for stop place categorization," J. Acoust. Soc. Am. **90**, 1309–1325.
Tan, Q., and Carney, L. H. (**2006**). "Predictions of formant-frequency discrimination in noise based on model auditory-nerve responses," J. Acoust. Soc. Am. **120**(3), 1435–1445.
Tyler, R. S., Hall, J. W., Glasberg, B. R., Moore, B. C. J., and Patterson, R. D. (**1984**). "Auditory filter asymmetry in the hearing impaired," J. Acoust. Soc. Am. **76**, 1363–1368 (1984).
Wolfe, J., John, A., Schafer, E., Nyffeler, M., Boretzki, M., Caraway, T., and Hudson, M. (**2011**). "Long-term effects of non-linear frequency compression for children with moderate hearing loss," Int. J. Audiol. **50**, 396–404.
Wolfe, J., John, A., Schafer, E., Nyffeler, M., Boretzki, M., and Caraway, T. (**2010**). "Evaluation of nonlinear frequency compression for school-age children with moderate to moderately severe hearing loss," J. Am. Acad. Audiol. **21**, 618–628.
Zilany, M., Bruce, I., and Carney, L. (**2014**). "Updated parameters and expanded simulation options for a model of the auditory periphery," J. Acoust. Soc. Am. **135**, 283–286.