Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Psychological and Physiological Acoustics
Session 4pPP: Computational Modeling of Sensorineural Hearing Loss: Models and Applications

4pPP4. Neural-scaled entropy as a model of information for speech perception

Joshua M. Alexander* and Varsha Hariram

***Corresponding author's address: Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana 47907, alexan14@purdue.edu**

Neural-Scaled Entropy (NSE) is an objective metric used to quantify 'information' available in speech consequent hearing loss, hearing aid signal processing, and distortion from various environmental factors. One pursuit is to use NSE to find optimum hearing aid settings that maximize speech perception. Inspired by the Cochlear-Scaled Entropy model [Stilp et al., 2010, J. Acoust. Soc. Am., 2112-2126], NSE uses the neural spike output at the inner hair cell synapse of an auditory nerve model [Zilany et al. 2009, J. Acoust. Soc. Am., 126, 2390-2412]. Probability of spike output from fibers sampled at equidistant places along the model cochlea is computed for short duration time frames. Potential information is estimated by using the Kullback-Liebler Divergence to describe how the pattern of neural firing at each frame differs from preceding frames in an auto-regressive manner. NSE was tested using nonsense syllables from various perceptual studies that included different signal processing schemes and was compared to performance for different vowel-defining parameters, consonant features, and talker gender. NSE has potential to serve as a model predictor of speech perception, and to capture the effects of sensorineural hearing loss beyond simple filter broadening. [Supported by NIDCD RC1DC010601]

Published by the Acoustical Society of America through the American Institute of Physics

Received 23 Jan 2013; published 2 Jun 2013

INTRODUCTION

Motivation for yet another Computational Model of Speech Processing

A variety of signal processing algorithms have been developed for hearing aids in an attempt to restore speech information to individuals with degraded sensory processing. The purpose of this project is to develop an objective measure that is sensitive to the amount of information in the altered speech signal in order to better understand the perceptual effects of these hearing aid processing strategies and to make predictions for future strategies. Of particular interest and where we began our investigation is a class of signal processing algorithms known as frequency lowering. We started here because to date very little is understood about how these processed signals are coded in the auditory periphery (normal or impaired) or even what constitutes 'information' as they are encoded further up the auditory system.

Individuals with hearing impairment have difficulty perceiving high-frequency speech information with conventional amplification. As the severity of loss increases, so does the range of frequencies affected. Frequency lowering algorithms attempt to recode this inaudible information by shifting high-frequency speech cues to low and mid frequency regions where sensory processing is relatively intact. This project initially focused the most popular form of frequency lowering – nonlinear frequency compression (NFC) because, unlike other techniques, the nuances associated with the algorithm, such as, when and where lowering occurs do not depend on the characteristics of the input signal. Therefore, this technique can be more easily replicated in the lab, resulting in a significant amount of perceptual data. With NFC, the input spectrum is functionally divided into two parts at a nominal frequency designated as the 'start frequency.' The spectrum below the start frequency is unaltered whereas the spectrum above the start frequency is shifted down in frequency by an amount that depends on the frequency distance of the individual spectral components from the start frequency. Spectral components closest to the start frequency undergo the least frequency shifting on a linear scale and those furthest from the start frequency undergo the greatest frequency shifting (Simpson, Hersbach, & McDermott, 2005). The exact amount frequency shifting and the subsequent reduction in bandwidth above the start frequency depends on the nominal 'compression ratio' with greater ratios corresponding greater frequency shifts (Figure 1). In fact, because NFC is logarithmic, the nominal compression ratio is about equal to the reduction in spectral resolution in terms of auditory filters. For example, a 2.0:1 compression ratio means that information that would normally span two auditory filters in the un-impaired ear before processing would only span one auditory filter after processing.

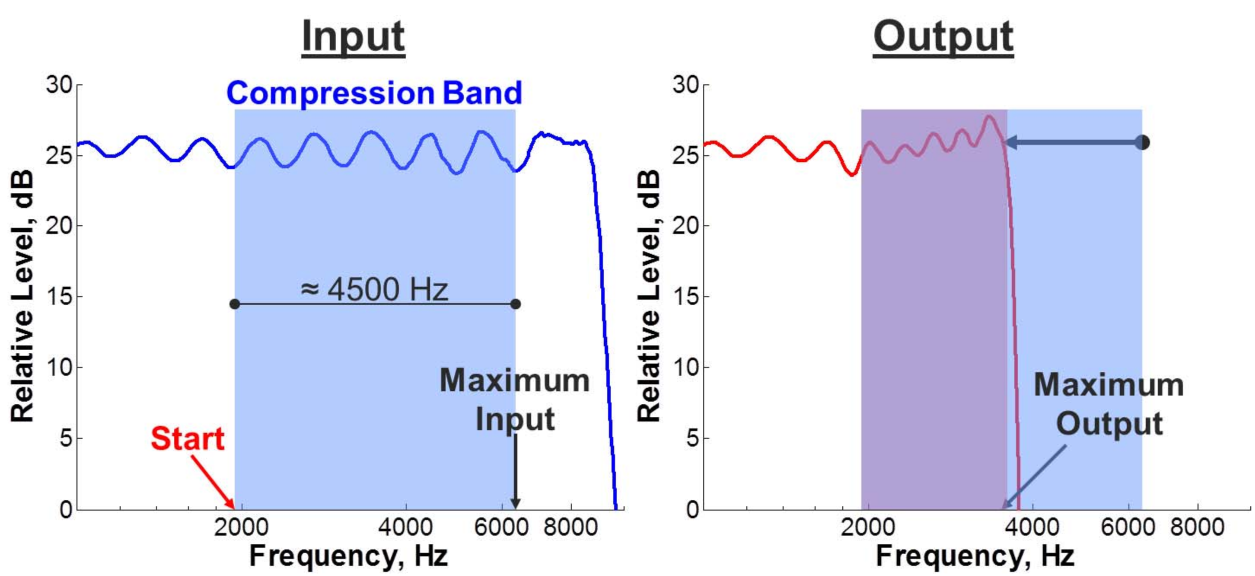

FIGURE 1. The figure on the left shows the spectrum a rippled noise with spectral peaks at the 1/3 octave frequencies before processing with nonlinear frequency compression (NFC). In this example, frequencies in a 4500-Hz band from stretching 1800 Hz (the 'start frequency') are shifted down in frequency. The figure on the right shows the same signal after processing with NFC. Only the frequencies in the compression band (blue-shaded region) are shifted (red-shaded region); frequencies below are unaltered by NFC and frequencies above are filtered off as a byproduct of the processing.

It follows that the integrity of information following recoding with NFC is dictated by the start frequency and the amount of compression. It also depends on the characteristics of individual speech sounds. For example, prior perceptual studies (Alexander, 2012) found that high-frequency speech content can undergo substantial frequency compression and still contribute to intelligibility of fricatives. This is likely because frication information is spectrally diffuse and is therefore less dependent on precise frequency content. In contrast, alternation of the primary formants, which originate at low to mid frequencies (Figure 2), is not as perceptually forgiving and serious disruptions in vowel identification, for example, can occur as start frequency decreases. In other words, where the compression occurs appears to be much more important that the amount of compression *per se*. What is needed is an index that quantifies the information in speech in a way that is sensitive to the signal alterations introduced by algorithms, like NFC, and by sensorineural hearing impairment.

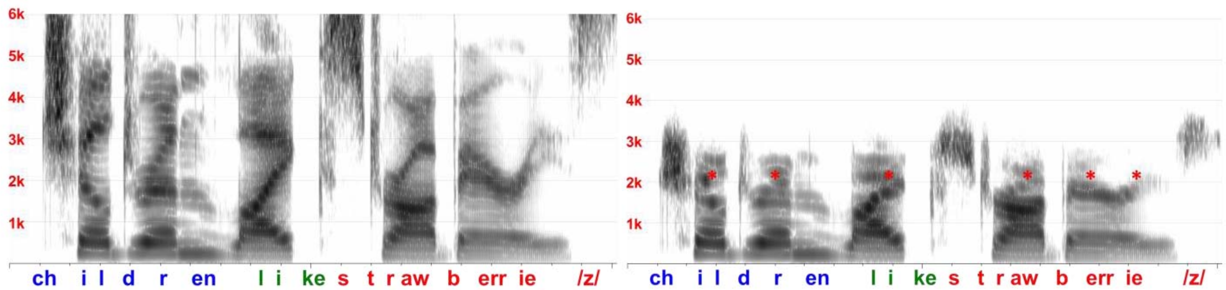

FIGURE 2. The figure on the left shows the spectrogram of the sentence “children like strawberries” as spoken by a female talker before processing with NFC. The figure on the right shows the same sentence after processing with NFC, whereby frequencies from 1.5-6.0 kHz are shifted down in frequency. It is hypothesized that because the spectral energy of the fricatives and stop busts is diffuse that they lose less information following NFC. In contrast, the spectral energy of the vowels and semivowels is concentrated in narrow bands at specific formant frequencies; therefore, they are more prone to information loss following NFC. For example, formant transitions become flattened following processing with a low start frequency as denoted by the asterisks on the figure on the right.

Limitations of Previous Computational Models

Several measures have been developed to predict speech intelligibility following hearing impairment and amplification. For example, the Speech Intelligibility Index, SII (ANSI, 2007) is based on the assumption that intelligibility of a speech signal depends on the total information within individual frequency-weighted bands, whereby information is determined primarily by the audibility of the dynamic range of speech in each band. Decreases in information, presumably associated with cochlear processing (e.g., broadened tuning at high presentation levels, upward spread of masking, and severe inner hair cell loss or cochlear ‘dead regions’), have been built into the model using correction factors that deweight the information in the appropriate bands. At least one proposal (Bentler, Cole, & Wu, 2011) has been made to modify the SII in order to predict intelligibility for users of hearing aids with NFC by taking the frequency importance weight from the speech before lowering and then assigning audibility based on the frequency region where it is moved. The problem with these models is that they do not account for the fact that frequency lowering introduces distortion and that the amount of this distortion will vary depending on the input frequencies and the frequency lowering parameters.

Other measures such the Speech Transmission Index (STI), which is based on the modulation transfer function, explicitly try to predict speech intelligibility amidst temporal and non-linear distortions (Bondy, Bruce, Becker, & Haykin, 2004). Despite this difference from SII, the STI is not particularly designed to make predictions based on underlying auditory mechanisms such as auditory filter broadening and has been shown to be an inferior predictor speech information in the presence of certain temporal distortions (Stilp, Kiefte, Alexander, & Kluender, 2010). Therefore, for our purposes, these metrics are likely to fail without some sort of modification (Hines & Harte, 2010).

Cochlear Scaled Entropy (CSE)

Cochlear Scaled Entropy (CSE) offers an attractive starting point for our new model. CSE uses concepts from Shannon information theory, which essentially states that information transfer occurs when uncertainty is reduced (in lay terms, 'when the unknown becomes known') so that the more unpredictable a system is, the greater the potential for information transfer (Shannon, 1948). When formally quantified in terms of bits, the measure of uncertainty is 'entropy.' Speech is a more or less orderly signal with a lot of redundancy across time and frequency. Where signal redundancy is low, it is less predictable and uncertainty/entropy are high. Consequently, entropy can be a direct measure of the *potential* speech information that can be conveyed through a system. 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 wishes to explore. For speech, this could be at the level of linguistic units such as syllables or whole words (Shannon, 1951), at the level of the auditory periphery (e.g., Stilp *et al*., 2010), or somewhere in between the two extremes. In other words, decisions need to be made about the level of processing one wishes to explore and how one chooses to represent the processing at that 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 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 simplified time-frequency representation of the speech. Time is divided into non-overlapping segments (most often 16 ms) 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 & 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 only to changes (that which is unpredictable) 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 (or unpredictable) from one or more preceding slices (Stilp & Kluender, 2010). Dissimilarity is estimated by the Euclidean distance between successive spectral slices, or across a running average of spectral slices (most often, 80 ms or 5 slices), and is a proxy for the information-theoretic measure of entropy.

CSE has been successfully used to explain intelligibility of speech under a variety of temporal distortions, including conditions in which measures based on the modulation transfer function fail (Stilp *et al*., 2010). In addition, it has been shown that when groups of spectral slices that have been identified as high CSE are replaced by speech-shaped noise that speech intelligibility is significantly degraded and when low CSE spectral slices are replaced that speech intelligibility is largely maintained (Stilp & Kluender, 2010). This same study showed that vowels and semi-vowels had greater entropy compared to consonants, which is consistent with previous findings that demonstrated a superiority of vowels over consonants for speech intelligibility (Kewley-Port, Burkle, & Lee, 2007). This finding has important implications for how the proposed model, an extension of CSE, is predicted to behave with frequency-lowered speech. That is, because CSE utilizes a quasi-logarithmic scale it gives more weight to lower frequencies, which often 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 frequency lowering, it should have a greater impact on entropy compared to a reduction in the bandwidth of frication, which tends to naturally fall in relatively broader high-frequency filters. This influence on entropy would be consistent with perceptual findings (Alexander, 2012).

Moving Beyond the Cochlea

As the name suggests, CSE is a measure of change at the cochlear level of auditory processing. While parsimonious, one cannot ignore strong evidence suggesting that speech perception is also shaped by processes further up the auditory pathway. Adaptation and suppression are two such phenomena present at the level of the auditory nerve. 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 auditory nerve discharge peaks that correspond to spectro-temporal regions rich in phonetic. In other words, adaptation may improve the neural representation of spectral change (entropy) between following speech segments (Kluender & Alexander, 2007).

Therefore, the aim of this project is to use the framework established by CSE and incorporate influential processes from further up the auditory system to develop a measure of speech intelligibility that is sensitive to information at the neural level. By using the Zilany *et al*. (2009) phenomenological model of the cat auditory nerve (AN) to describe the time-frequency representation of speech at the level of the inner hair cell-auditory nerve synapse, the new entropy measure attempts to capture the effects of adaptation, suppression, and other nonlinearities; thereby, allowing one to model the effects of sensorineural hearing loss beyond audibility and simple filter broadening.

METHODS

Neural Time-Frequency Representation of Speech

The Zilany *et al*. (2009) AN model consists of several modules representing the peripheral auditory structures starting at the middle ear all the way to the auditory nerve. Studies have shown that this and previous versions of the model can be used to predict AN responses for a wide variety of stimuli spanning the dynamic range of hearing for both normal and impaired ears (Bandopadhyay & Young, 2004; Tan & Carney, 2006; Hines & Harte, 2010). An advantage for our purposes is that it includes various aspects of peripheral auditory impairment including inner hair cell (IHC) and outer hair cell (OHC) dysfunction, and consequences in terms of loss of audibility, loudness growth and broadened tuning. In order to account for sensorineural hearing impairment, the constants C_{IHC} and C_{OHC} can be varied to control the scaling factors for IHCs and OHCs respectively. The specific values C_{IHC} and C_{OHC} vary depending on the CF and the amount of audiometric hearing loss, with 0 corresponding to total impairment and 1 corresponding to no impairment. Normative data is used within the model to generate these values for a given audiogram. By default, the model assumes that the proportion of audiometric loss attributed to OHC and IHC impairment is 2/3 and 1/3, respectively (Bruce, Sachs, & Young, 2003; Zilany & Bruce, 2007; Moore, Glasberg, & Vickers, 1999).

Before processing with the AN model, audio files were passed were through a transfer function for a BeyerDynamic DT150 headphone in order to make the input to the model comparable to the signals heard by participants in behavioral studies (Figure 3).

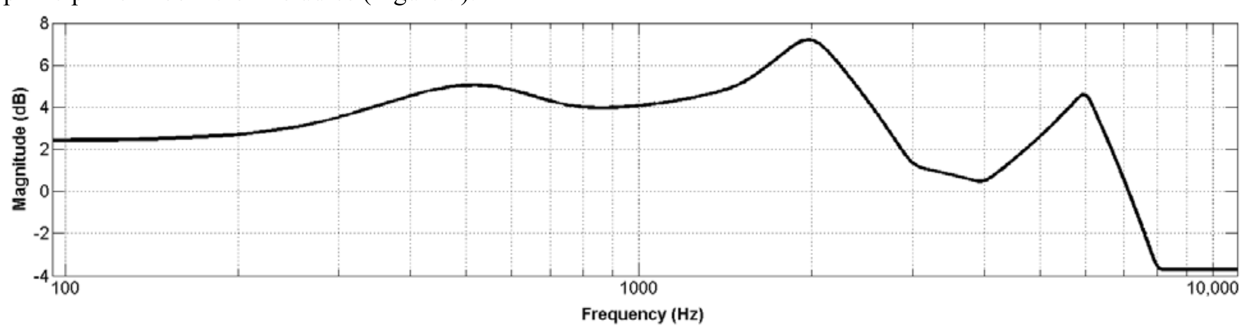

FIGURE 3. The 176-tap finite impulse response filter that was used to shape the audio files to simulate the frequency shaping of the headphone output by the external ear.

The filtered audio files were then processed by the AN model and output at the IHC-AN synapse in spikes/second 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: $F=A(10^{ax} – k)$, where F=frequency in Hz, A=165.4, a=0.06, k=0.88 and x=distance in mm. The simulated presentation level, fiber spontaneous rate (SR), and values for C_{IHC} and C_{OHC} depended on the condition being tested. The final step to generating the neural time-frequency representation, the neurogram, involved averaging the simulated neural response over manageable time frames (e.g., 5-15 ms). Figure 4 is an example of a neurogram from un-impaired medium SR fibers using 5-ms frames for the speech sound /aʃa/ as spoken by a male talker at 60 dB SPL mixed with speech-shaped noise at 10 dB SNR. Spike rate is depicted on the Z-axis for a given CF (Y-axis) and time frame (X-axis). The initial and final vowel formants can easily be visualized in low frequencies and the frication noise associated with medial /ʃ/ can be seen in the high frequencies.

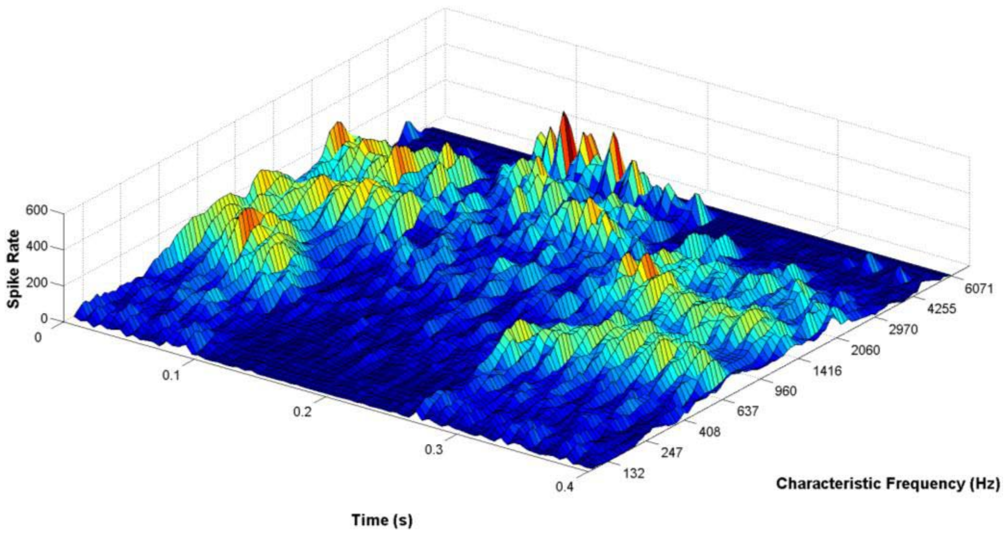

FIGURE 4. A 3-D neurogram (Y-axis, characteristic frequency in Hz; X-axis, time in secs; Z-axis, spike rate in spikes/sec) from medium spontaneous rate fibers for the speech sound /aʃa/ produced by a male talker mixed with speech-shaped noise.

Neural-Scaled Entropy (NSE)

To compare how the cochleotopic pattern of neural firing changes across time frames, Euclidean distance can be computed just as with CSE, but in terms of spike rate. However, for a more 'pure' estimate of entropy, the Kullback-Liebler Divergence (KLD) can be computed, which quantifies how much one probability distribution differs from another (Johnson, Gruner, Baggerly, & Seshagiri, 2001; Bandopadhyay & 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 (Figure 5a). To compute NSE, KLD is measured for each time frame when predicted by a set number of prior frames (e.g., a 50-ms window of ten 5-ms frames) in an auto-regressive manner (Figure 5b).

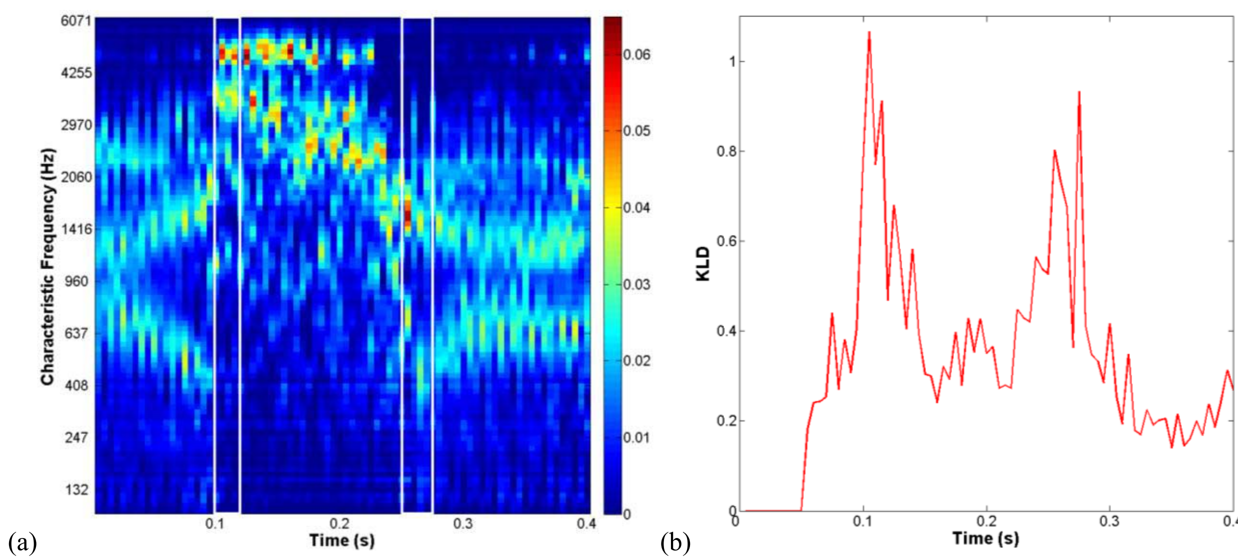

FIGURE 5. (a) A 2-D representation of the neurogram in Fig. 4 but in terms of relative probability (see color bar), that is, normalized spike rate for each time frame. The white boxes in (a) indicate regions with the greatest KLD when averaged over 50-ms windows (ten 5-ms frames) as shown in (b). NSE is not computed until after the first averaging window.

With the representation of neural spikes as a series of probability distributions, one can conceptualize the neural transfer of information as a process by which higher auditory structures must 'learn' which CF generated each spike it receives (*i.e.*, uncertainty reduction). 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 (KLD) will be very low. In addition, changes across a limited number fibers like a formant transition will have relatively greater entropy than random changes that span a wide range of fibers like frication because probabilities across CF change very little with noise. Finally, the greatest entropy might be expected for boundaries between vowels and consonants where the distribution (probability) of neural firing can change between a high-CF skew and a low-CF skew in a short period as shown in Figure 5.

PRELIMINARY FINDINGS

As a preliminary test of our methods, model data was compared to perceptual data from an experiment involving 14 listeners with mild to moderate sensorineural hearing loss who identified nonsense syllables that were mixed with speech-shaped noise at 10 dB SNR and processed using a variety of NFC parameters (Alexander, 2012). Stimuli in the conditions below were subjected to a final stage of multichannel wide dynamic range compression to compensate for the loss of audibility and were then low-pass filtered at about 3.3 kHz to simulate a severe (uncompensated) high-frequency hearing loss. The control condition was processed with the low-pass filtering, but was not processed with NFC. In a 2x3 design, two start frequencies (about 1.6 and 2.2 kHz) were crossed with 3 input bandwidths (about 5, 7, and 9 kHz). That is, the speech spectrum ranging from the start frequency to the upper frequency in the input bandwidth were nonlinearly compressed so that the latter was lowered to 3.3 kHz, the maximum amplified frequency. For a fixed start frequency, increases in audible input bandwidth were accomplished by increases in the compression ratio. Shown in the second column of Table 1 are the speech intelligibility results for stop consonants that were produced in three different vowel-consonant-vowel contexts (/a/, /i/, and /u/) by 2 male and 2 female talkers. As can be seen, under these conditions average performance for all the NFC parameters was worse than the control, with the worst performance for those conditions that altered formant frequencies the most (a 1.6 kHz start frequency combined with high compression ratios for the 7 and 9 kHz input BWs). With the 2.2 kHz start frequency, the amount of compression made very little difference, likely because the critical formant frequencies were either unaltered or altered to a lesser extent.

TABLE 1. Displayed are outcomes for different conditions in which stop consonants were processed with nonlinear frequency compression using 2 start frequencies and 3 compression ratios that corresponded to a progressively increasing audible input bandwidth (BW). Perceptual outcomes are shown in the second column and model data in the remaining columns (see text). The last row indicates the amount of variance accounted for by each model (R-squared).

Condition	Perceptual	CSE	NSE	NSE (KLD)
1.6 kHz Start, 5kHz BW	58.3%	0.78	20.1	0.048
1.6 kHz Start , 7kHz BW	53.1%	0.68	17.9	0.043
1.6 kHz Start, 9kHz BW	50.1%	0.65	17.2	0.039
2.2 kHz Start, 5kHz BW	59.7%	0.74	20.0	0.046
2.2 KHz Start, 7kHz BW	60.0%	0.70	19.2	0.044
2.2 KHz Start, 9kHz BW	58.9%	0.68	19.0	0.044
Low-Pass Control 3.3 kHz	63.7%	0.94	21.6	0.047
R-squared with Perceptual Data		0.55	0.86	0.67

Three models were compared to the perceptual data. For each model, the best matches to the data, which are shown in Table 1, were obtained when speech-in-quiet was processed and when it was assumed that cochlear function was un-impaired. As of now, including noise and impairment in the models represents a limitation of our current methods. Results from the CSE model are shown in column 3. CSE was obtained using the same code as Stilp & Kluender (2010), in which Euclidean distance between successive 16-ms spectral slices was measured. For comparison, NSE in terms of Euclidean distance (denoted 'NSE' in column 4) and KLD (column 5) were also computed between successive 16-ms time frames. Fibers with CFs corresponding to 0.25-mm samples along the cochlear partition from about 0.1 to 3.3 kHz (78 total) were included in the computations. NSE was carried out separately for fibers with low, medium and high SRs, as prescribed in the Zilany *et al*. (2009) AN model. The final estimate of NSE was a weighted average across fiber types, using 0.61, 0.23, and 0.16 as weights for the high, medium, and low SR fibers, respectively, which corresponds to the estimate of their distribution in the cat cochlea (Liberman, 1978; Zilany *et al*.M, 2009).

For all the models, R-squared for just the three conditions with the 1.6 kHz start frequency was ≥ 0.97, meaning that they accurately described the loss of information brought about by increased amounts of low-frequency compression. As shown in Table 1, a distinction between the models occurs when trying to account for the differences in performance for the 2.2 kHz start frequency. CSE, which models auditory filter tuning only, could account a little more than half of the variance in the perceptual data, while NSE (KLD) explained a slightly greater proportion, and NSE using Euclidean distance explained the most (over 86%). One reason why the Euclidean distance of neural responses does better than KLD in explaining the data for this example might be because KLD implicitly normalizes firing rate across time frames, thereby losing potential information conveyed by differences in firing rate across time. It remains to be seen whether this superiority holds up in other conditions and for analyses involving other phonemes.

DISCUSSION

The purpose of this project was to develop an objective measure, neural-scaled entropy (NSE), that is sensitive to the amount of information in the altered speech signal in order to better understand the perceptual effects of different hearing aid processing strategies and to make predictions for future strategies. What was presented here represents only a start that involved one specific type of frequency lowering. We have perceptual data for other phonemes and NFC conditions that need to be modeled (Alexander, 2012). In addition, it remains to be seen how closely the models can explain observed differences associated with talker gender, manner of articulation, and different vowel defining parameters, such as vowel context for consonant identification and vowel height and tongue advancement for vowel identification. Several details about the computation of NSE have yet to be explored, including the effect of different duration time frames and the number the recursive comparisons. Furthermore, it is not yet apparent how NSE across the different SRs or across time should be considered. Averaging is only one option; other options included a more judicious selection of SRs based on presentation level or simply selecting the fibers and/or time frames with the greatest NSE, as with a winner-take-all strategy. Finally, as mentioned earlier, we have yet to explore why we had difficulty when generating models for noise-contaminated speech and for amplified speech presented to a model of the impaired auditory periphery. Previous considerations might be useful in this regard.

ACKNOWLEDGMENTS

We thank Michael Heinz and Keith Kluender for their helpful discussions during the ongoing development of this model. Initial funding provided by NIDCD RC1DC010601.

REFERENCES

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.

ANSI. (**2007**). *ANSI S3.5-1997 (R2007), Methods for Calculation of the Speech Intelligibility Index.* New York: American National Standards Institute.

Bandopadhyay, S., & Young, E. D. (**2004**). Discrimination of voiced stop consonants based on auditory nerve discharges. *The Journal of Neuroscience,* ***24(2)***, 531-541.

Bentler, R., Cole, W., & 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.

Bondy, J., Bruce, I. C., Becker, S., & Haykin, S. (**2004**). Predicting speech intelligibility from a population of neurons. *NIPS 2003 Conference Proceedings: Advances in Neural Information Processing Systems,* ***16***. Cambridge, MA: MIT.

Bruce, I. C., Sachs, M. B., & Young, E. D. (**2003**). An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses. *Journal of the Acoustical Society of America, **113***, 369–388.

Delgutte, B. (**2002**). Auditory Neural Processing of Speech. In W. Hardcastle, & J. Laver, *The Handbook of Phonetic Sceinces* (pp. 507-538). Oxford: Blackwell.

Glasberg, B. R., & Moore, B. C. (**1990**). Derivation of auditory filter shapes from notched-noise data. *Hearing Research, **47***, 103-138.

Greenwood, D. D. (**1990**). A cochlear frequency-position function for several species-29 years later. *Journal of Acoustical Society of America, **87(6)***, 2592-2605.

Hines, A., & Harte, N. (**2010**). Speech intelligibility from image processing. *Speech Communication, **52***, 736-752.

Johnson, D. H., Gruner, C. M., Baggerly, K., & Seshagiri, C. (**2001**). Information-Theoretic Analysis of Neural Coding. *Journal of Computational Neuroscience, **10***, 47-69.

Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (**2007**). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *Journal of the Acoustical Society of America*, ***122*** 2365–2375.

Kluender, K. R., & Alexander, J. M. (**2007**). Perception of Speech Sounds. In P. Dallos, & D. Oertel, *Handbook of the Senses: Audition* (pp. 1-28). London: Elsevier.

Kluender, K. R., Coady, J. A., & Kiefte, M. (**2003**). Sensitivity to change in perception of speech. *Speech Communication, **41***, 59–69.

Liberman, M. C. (**1978**). Auditory-nerve response from cats raised in a low-noise chamber. *Journal of the Acoustical Society of America, **63***, 442-455.

Moore, B., Glasberg, B. R., & Vickers, D. A. (**1999**). Further evaluation of a model of loudness perception applied to cochlear hearing loss. *Journal of the Acoustical Society of America, **106***, 898-907.

Shannon, C. E. (**1948**). A mathematical theory of communication. *Bell Systems Technical Journal, **27***, 379–423 and 623–656.

Shannon, C. E. (**1951**). Prediction and entropy of printed English. *Bell Systems Technical Journal, **30***, 50-64.

Simpson, A., Hersbach, A. A., & McDermott, H. J. (**2005**). Improvements in speech perception with an experimental nonlinear frequency compression hearing device. *International Journal of Audiology, **44***, 281-92.

Stilp, C. E., & Kluender, K. R. (**2010**). Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility. *Proceedings of the National Academy of Sciences, **107(27)***, 12387–12392.

Stilp, C. S., Kiefte, M., Alexander, J. M., & Kluender, K. R. (**2010**). Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentences. *Journal of Acoustical Society of America, **128(4)***, 2112–2126.

Tan, Q., & Carney, L. H. (**2006**). Predictions of formant-frequency discrimination in noise based on model auditory-nerve responses. *Journal of Acoustical Society of America, **120(3)***, 1435–1445.

Zilany, M. A., Bruce, I. C., Nelson, P. C., & Carney, L. H. (**2009**). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics. *Journal of Acoustical Society of America, **126(5)***, 2390–2412.

Zilany, M. S., & Bruce, I. C. (**2007**). Representation of the vowel in normal and impaired auditory nerve fibers: Model predictions of responses in cats. *Journal of the Acoustical Society of America, **122***, 402–417.