Representations of fricatives in sub-cortical model responses: comparisons with human perception

Yasmeen Hamza[1]

University of Rochester, Department of Biomedical Engineering

Afagh Farhadi

University of Rochester, Department of Electrical and Computer Engineering

Douglas M. Schwarz

University of Rochester, Depts. of Neuroscience and Biomedical Engineering

Joyce M. McDonough

University of Rochester, Department of Linguistics

Laurel H. Carney

University of Rochester, Depts. of Biomedical Engineering, Neuroscience, and Electrical and Computer Engineering

Date of Submission: 09/30/2022

Keywords: Fricatives; speech perception; hidden hearing loss; computational model, auditory processing.

Running Title: Representations of fricatives in sub-cortical model

[1] Author to whom correspondence should be addressed.
Current Affiliation: Florida State University, School of Communication Science and Disorders, College of Communication and Information, Tallahassee, Florida, USA.
Electronic mail: yhamza@fsu.edu
The work was continued while affiliated at University of California Irvine (2019-2022) and Florida State University (2022-)

Portions of this work were presented in "Can Auditory-Nerve and Inferior Colliculus Models Explain Perceptual Confusions for Fricatives? Association for Research in Otolaryngology 43rd Midwinter meeting, San Jose, CA, USA, 2020."

This paper is part of a special issue on Perception and Production of Sounds in the High-Frequency Range of Human Speech.

ABSTRACT

Fricatives are obstruent phonemes made by constricting airflow in the vocal tract, producing turbulence across the constriction site or downstream from the constriction. Fricatives exhibit significant intra/inter-subject and contextual variability. Yet fricatives are perceived with high accuracy. The current study investigated modeled neural responses to fricatives in the auditory nerve (AN) and inferior colliculus (IC), with the hypothesis that response profiles across populations of neurons provide robust correlates to perception. Stimuli were 270 intervocalic fricatives (10 speakers x 9 fricatives x 3 utterances). Computational model response profiles had 50 characteristic frequencies, log-spaced from 125 to 8 or 20 kHz. Confusion matrices generated by KNN subspace classifiers used the average rates for the 50 characteristic frequencies as the features and compared them with published behavioral data. The modeled AN and IC neural responses provided better predictions of perceptual accuracy than the stimulus spectra, with IC showing better accuracy than AN. Perceptual fricative accuracy was explained by modeled neural response profiles, whereas confusions were only partially explained. Extended frequencies improved accuracy based on the model IC, corroborating the importance of extended high frequencies in speech perception. Future electrophysiological studies from IC cells and neural response simulations of hearing-impaired ears could guide the development of processing strategies.

I. INTRODUCTION:

A. Fricatives

Fricatives are the set of obstruent sound contrasts made by constricting airflow in the vocal tract, which produces turbulence across the constriction or a site downstream from the constriction (Ladefoged, 1971; Ladefoged and Maddieson, 1996; Maddieson and Disner, 1984). Fricatives occur at every place of articulation, the only phonemes to do so, and they may exhibit voicing contrasts[2]. The IPA lists twelve distinct places of articulation for fricatives, from labial to pharyngeal and glottal. Out of the 315 languages in Maddieson and Disner's (1984) original UPSID database[3], 296 have at least one fricative, with four fricatives as the mean number. The coronal (dental through palatal) fricatives are among the most common sounds found in phoneme inventories (Maddieson and Disner, 1984). Phonemic inventories exhibit preferences for voiceless fricatives and avoid voicing distinctions, but the distribution of voicing in fricatives is not symmetric. The most common fricative contrast is the dental or alveolar fricative, /s/, followed by /ʃ/ and /f/. The next most common are the voiced /z/, /ʒ/, and the voiceless velar fricative, /x/. Paired voicing distinctions do not start to appear until a language has four fricatives, but voicing distinctions do not always appear in pairs. While the /z/ and /ʒ/ don't occur in the database without their voiceless counterparts, the voiced labial /β/, labiovelar /v/, interdental /ð/ and velar /ɣ/ appear with regularity without voiceless counterparts (Maddieson 1984:44-57).

[2]Maddieson (1984) excludes the glottal fricative /h/ from the class of fricatives.

[3]UPSID (UCLA Phonological Segment Inventory Database) is a DOS based database of contrastive segment inventories produced by Ian Maddieson (1984). The original sample was 317 languages but was updated to 451 (Maddieson and Precoda 1990). An online version of the database is currently at http://www.phonetik.uni-frankfurt.de/upsid_info.html .

While intensity is difficult to define or measure in speech, an early study of intensity by Strevens (1960) as muscular effort and air pressure demonstrated that there is little correlation between intensity and high frequency energy in fricatives. In Ladefoged and Maddieson's (1996) landmark book (*Sounds of the Worlds Languages*) they note that, in comparison to other sounds, significant demands are involved in the articulation of fricative sounds, in the precision needed to shape the particular airstream channel and in maintaining the target shape over the duration of the sound.

Most of the work on fricative sounds has been based on acoustics and aerodynamics, supported by perception studies (Chodroff and Wilson, 2020; Hughes and Halle, 1956; Jassem, 1979; Proctor et al., 2010; Shadle, 1985; Shadle et al., 1992; Strevens, 1960). Less focus has been placed on their neural coding in the auditory pathway. The auditory system is critical because of the difficulty of identifying the acoustic properties which may serve as cues to the contrasts (Jongman et al., 2000; Narayanan et al., 1995). It has been long observed that, despite the complexity of the articulation, there are great differences in the acoustic signatures of a given fricative among different speakers. Shadle (1985) in particular has pointed out that fricatives may be best identified by their transitions in and out of the segment, their contexts, rather than in the articulation of the sound. In a cross linguistic study of voiceless fricatives from field data of seven languages, Gordon et al. (2002) found that spectral shape and center of gravity were important in distinguishing most of the fricatives in the study. They also noted that in cases where the fricatives had similar spectra and centers of gravity, the surrounding formant frequencies served to differentiate the fricatives.

Fricatives exhibit considerable intra/inter-subject variability (Jongman et al., 2000; Shadle, 1985; Shadle, 1990), as well as contextual variability (Narayanan et al., 1995). Previous research has attempted to identify temporal (duration and amplitude), and spectral properties that are responsible for the correct identification of fricatives (Silbert and De Jong, 2008). Acoustically,

fricatives are characterized by non-periodic energy at high frequencies, especially sibilant fricatives. The most common ways used to measure fricative sounds are the center and degree of dispersion of the fricative energy: skewness, center of gravity, and dispersion. Apart from the voicing-induced low-frequency spectral peak, voicing affects the spectra of fricatives primarily by reducing the amplitude of the noisy components (Crystal and House, 1988; Jesus and Shadle, 2002; Klatt, 1976). Unlike voicing, the place of articulation affects spectral properties of fricatives significantly but is also subject to a great deal of individual and contextual variability that does not appear to impede perception (Chodroff and Wilson, 2020; 2022; Hughes and Halle, 1956; Jesus and Shadle, 2002; Jongman et al., 2000; Shadle, 1990; Shadle et al., 1992).

Acoustic differences between sibilant (/s/, /z/, /ʃ/, /ʒ/) and non-sibilant (all others) fricatives, including spectrum, amplitude, and duration, have been previously demonstrated (Behrens and Blumstein, 1988; Evers et al., 1998; Hughes and Halle, 1956; Shadle, 1990; Strevens, 1960). Sibilant fricatives involve a noise source that is downstream of constriction, thus their spectrum is shaped by the resonant frequencies of the cavity anterior to the constriction (Stevens, 1998), whereas non-sibilant fricatives channel turbulence only at the site of constriction (Catford, 1977). This difference results in sibilant-fricatives typically exhibiting a midfrequency spectral peak, as a result of the intense high-frequency turbulence caused by the airstream hitting the teeth, as opposed to non-sibilant fricatives that have a flat spectrum (Jongman et al., 2000; Shadle, 1990; Shadle et al., 1992) . Additionally, sibilant fricatives are of relatively higher amplitude than non-sibilants.

The spectral differences described above have been demonstrated through the use of statistical techniques such as spectral moments (Forrest et al., 1988; Jesus and Shadle, 2002; Jongman et al., 2000); however, such temporal and spectral features are not independent and are commonly

variable across subjects and contexts (Chodroff and Wilson, 2022; Jongman et al., 2000; Narayanan et al., 1995; Shadle, 1990). For example, it is common that phonologically voiced fricatives undergo substantial devoicing in certain contexts (Haggard, 1978; Stevens et al., 1992) and phrase final segments tend to be lengthened (Klatt, 1976). Despite this variability, listeners have been shown to be accurate in the perception of these contrasts (Cutler et al., 2004; Gallun and Souza, 2008; Pisoni and Luce, 1986; Woods et al., 2010). Nearly all accounts point to one thing: the importance of the auditory system in understanding fricative contrasts. A detailed understanding of the neural representation of these sounds, and its transformation from the auditory nerve to the midbrain will help bridge the gap between production and perception.

While much phonetic work has focused on the articulation and acoustic analysis of fricatives, minimal investigation of their representation in the auditory system has been done. The broad goal of the current study is to investigate the modeled neural responses of fricatives in the auditory-nerve (AN) and midbrain (central nucleus of inferior colliculus, IC) responses, to aid in understanding how fricatives are perceptually salient despite their great variability. The hypothesis is that response profiles across populations of neurons provide robust correlates to perception.

B. Auditory-nerve and inferior colliculus modeling

The relationship between acoustics and neural encoding is complex due to strong non-linearities in the auditory periphery. Computational models for auditory neurons provide a strategy to describe these transformations. Auditory-nerve (AN) responses are affected by basilar membrane compression (Rhode, 1971), synchrony capture (Deng and Geisler, 1987; Miller et al., 1997; Young and Sachs, 1979), synaptic adaptation (Goutman and Glowatzki, 2007; Moser and Beutner, 2000; Raman et al., 1994; Westerman and Smith, 1984), and rate saturation (Sachs and Abbas,

1974; Yates, 1990; Yates et al., 1990). A computational model that includes all of these nonlinearities (Zilany et al., 2014; Zilany et al., 2009) was used in this study to characterize the responses of the population of AN fibers to fricatives.

Because of the highly nonlinear properties of the auditory periphery, the profile of discharge rates across the population of AN fibers is not a simple representation of the stimulus spectrum. The limitations of average-rate representations of AN fibers are often addressed by studying phase-locking of AN fibers to temporal fine structure (TFS) (e.g., Young and Sachs, 1979). However, not only is the TFS of fricative stimuli highly complex, the importance of high-frequency components in these stimuli limits the utility of phase-locking to the TFS to encode the full frequency spectrum of these sounds. Furthermore, the representation of the TFS of higher-frequency components of speech sounds is not likely to be critical in exciting auditory neurons at higher levels of the auditory pathway, such as the midbrain. However, relatively low-frequency fluctuations in AN responses, which are strongly elicited by noisy sounds such as fricatives, are important in exciting and suppressing neurons at higher levels of the auditory pathway (review: Carney, 2018).

It is interesting to consider the representations of complex sounds at the level of the midbrain, which is a nearly obligatory synapse for ascending projections from several brainstem nuclei (Schreiner and Winer, 2005). At the level of the IC, frequency tuning can be described by a characteristic frequency (CF, the frequency eliciting a response at the lowest sound level). Additionally, most neurons are tuned for amplitude modulation frequency, as described by modulation transfer functions (MTFs) and the best modulation frequency (BMF), the modulation frequency that elicits the greatest excitation or suppression in the MTF (Krishna and Semple, 2000 ; Joris et al., 2004; Nelson and Carney, 2007). Most IC neurons have BMFs in the range of voice

pitch (Langner, 1992). Most IC neurons are either excited (band-enhanced, BE) or suppressed (band-suppressed, BS) by amplitude modulations at the BMF (Kim et al., 2020; Kim et al., 2015). This study explored representations of fricatives in the average rates of populations of model BE and BS neurons (Carney et al., 2015; Nelson and Carney, 2004).

The current study utilized computational models for AN fibers and IC neurons to test the proposed hypothesis that perceptual accuracy for a fricative identification task involving stimuli with and without spectral degradation from Gallun and Souza (2008) could be better explained by the response profiles across populations of neurons than by the spectrum of the acoustical signal.

C. Hidden hearing loss

Hearing loss is conventionally defined as elevated thresholds within the frequency range of 125 Hz to 8 kHz (WHO, 2008). The limitation to 8 kHz in conventional pure-tone audiometry is based on the finding that much of the studied phonetic information is provided by frequencies below 6 kHz (Vitela et al., 2015). However, there is growing evidence suggesting that acoustic information in the higher-frequency regions affects speech intelligibility, particularly in noisy environments (Badri et al., 2011; Levy et al., 2015; Moore et al., 2017; Pienkowski, 2017; Zadeh et al., 2019). Hearing loss at the extended high frequencies is common with aging and noise exposure and may reflect cochlear synaptopathy, which occurs first at higher frequencies (Liberman et al., 2016). In the current study, we compared the accuracy of computational models for acoustic, AN fibers and IC neurons using information limited to 8 kHz vs. an extended frequency of 20 kHz. As previously stated, much of the phonetic information is limited to 6 kHz, therefore we hypothesized that accuracies of acoustic models would not differ with the limitation of frequencies. On the other hand, because high-frequency auditory neurons respond across a wide range of frequencies, in part

due to the low-frequency tails of the AN tuning curves (Kiang and Moxon, 1974), limiting the computational models to 8 kHz was expected to reduce accuracies of the AN and IC models, in line with the aforementioned evidence of the importance of high-frequency hearing in speech intelligibility.

II. METHODS

A. Stimuli

The recordings were done at 44.1kHz, with 16 bits, on a Marantz PDM661, using a Shure cardioid lavalier microphone. The stimuli were the four sets of English fricatives, contrasting in place of articulation and voicing: labio-dental (/f/, /v/), interdental (/θ/, /ð/), alveolar (/s/, /z/), palatal-alveolar (/ʃ/, /ʒ/) plus the glottal fricative (/h/) in an intervocalic aCa (vowel-fricative-vowel) context. Ten native English speakers (5 males and 5 females) were digitally recorded in a quiet room in the Linguistics Phonetics Lab while seated comfortably in the presence of the experimenter. The nine stimuli were read off a prepared list; participants were instructed to repeat each item three times, thus twenty-seven utterances were collected from each speaker.

The 270 utterances (10 speakers x 9 fricatives x 3 utterances) were segmented on Praat (Boersma and Weenink, 2016) and demarcation of the onset, offset and midpoint of the initial vowel and fricative was determined for each utterance by visual inspection of the spectrogram and waveform. Formant frequencies and voicing cues, when possible, were used as guides.

Each of the aCa tokens was then scaled so that the level of the initial /a/ mid-section (60 ms) was 65 dB SPL. This scaling allowed the preservation of the relative levels of sibilant vs. non-sibilant fricatives. In the current study we analyze the model neural responses to the mid-fricative segment (see Table1e in the supplement for mid-fricative RMS).

Guided by Gallun and Souza (2008), four spectrally-degraded, speech-test conditions were also created: fricatives were processed by digital filtering with one-, two-, four-, or eight-channels that spanned from 176 to 7168 Hz. Intermediate cutoff frequencies for the channels were logarithmically spaced. The filtered segments were then degraded by randomly multiplying each sample of the filter outputs by +1 or -1, before summing the signals together to create the processed stimuli (Schroeder, 1968).

B. Modeling

The response to each of the 270 aCa utterances in the average-rate profiles across populations of model AN and IC were simulated using computational models (Carney et al., 2015; Nelson and Carney, 2004; Zilany et al., 2014). Model response profiles were computed for 60-ms duration segments centered within each fricative. Model populations had 50 CFs, log-spaced from 125 to 20 kHz. To understand the importance of extended frequencies (beyond 8 kHz), models using the original speech condition (unprocessed, ie., no spectral degradation) were replicated while limiting the acoustical information and the model populations to 8 kHz, where model populations had 50 CFs, log-spaced from 125 to 8 kHz. Comparison of the confusion matrices of the spectra vs. neural-model responses in the extended (up to 20 kHz) vs. the conventional (8 kHz) frequency conditions would aid in understanding the potential contribution of the high-frequency channels.

The AN model used in the current study takes into consideration key non-linearities, previously mentioned, including compression, rate saturation, adaptation, and synchrony capture (Zilany et al., 2014). The model is based on the physiological responses of the AN in cat; here, we used a version with sharper peripheral tuning to represent the human ear (Ibrahim and Bruce, 2010), based on physiological and psychophysical measures (Shera et al., 2002).

The same-frequency inhibition-excitation (SFIE) model (Carney et al., 2015; Nelson and Carney, 2004) was used to model the responses of BE and BS IC neurons. The AN model provided the input for both types of IC model in the form of time-varying rate, convolved with functions representing excitatory or inhibitory postsynaptic responses, which differed for the two cell types (Carney et al., 2015). The postsynaptic potential time constants and delays were set to produce BE responses with a BMF of 100 Hz (Carney and McDonough, 2019). This BMF is near the center of the distribution of IC BMFs (Kim et al., 2020; Krishna and Semple, 2000; Nelson and Carney, 2007). Note that the modulation tuning of model and IC neurons is relatively broad ($Q \approx 0.5$-1, Nelson & Carney, 2007). The BS model receives an inhibitory input from the output of the BE model. The code used for the simulations is available here: Laurel Carney, Douglas Schwarz and Ava Elizabeth Giorgianni (2019-2022). UR_EAR Auditory-Nerve & Midbrain Simulation Code [computer program]. Retrieved on <date> from https://osf.io/6bsnt/.

C. Analysis

A k-nearest neighbors (KNN) classifier-based analysis was used to generate confusion matrices, from which the accuracy in predicting each fricative and the confusions between fricative contrasts were computed. Confusion matrices were constructed based on spectra, or on the response profiles for AN, BE, or BS models. Confusion matrices were then compared with published behavioral data for fricatives derived from the consonant-identification task based on the stimuli described above (Gallun and Souza, 2008). The set of features used for training and testing the classifier were the average rates for the 50 CFs, or spectral magnitudes at frequencies matching the log-spaced CFs. Model neural feature vectors were based on AN, BE, or BS responses.

The whole dataset (270 spectral/neural population responses) was divided using a cross-validation technique (20 folds) into training and testing subsets. The choice of 20 folds was based on the dataset size, and to reduce bias and variance in the models (Kohavi, 1995). In this case, the dataset was divided into 20 subsets, 19 were used to train the classifier and 1 was used to test the classifier. This was repeated 20 times, to use all possible training and testing combinations, and the average accuracy (percent-correct prediction) was calculated. Training the model was based on the feature vector of each of the data points in the training dataset, with a label identifying the class (fricative) of data. When the classifier was trained, the testing dataset was used to estimate the classes based on the training data.

Additionally, the classifier model was tested for overfitting by randomly splitting the dataset into training and test sets with size of 230 and 40 samples respectively. The classifier was trained using the training dataset and tested with both training and testing datasets for three trials. For all trials and for all type of responses, the difference between mean absolute error for training and testing sets was negligible (less than 2%).

The classifier type was a subspace, ensemble classifier that uses KNN learners. In the current study, the number of learners was 30 (spectral/neural responses per fricative), and the subspace dimension was 50, corresponding to the length of the feature vector. The MATLAB's Statistics and Machine Learning Toolbox was the platform used for the classifier analysis.

For the unprocessed conditions, training and testing were done using the unprocessed condition dataset, whereas for the processed conditions, training was done for the unprocessed condition and testing was done using the processed condition. The aim behind this strategy was to mimic the

process of the perceptual response in normal-hearing individuals, where degraded speech was compared to auditory signatures in memory to reach a decision.

III. RESULTS

Population average-rate responses to the unprocessed mid-fricative stimuli are shown as a function of model CF for sibilants (Fig. 1) and non-sibilants (Fig. 2). Additionally, the variation in the spectra and response profiles across the different processing conditions of a selected voiced sibilant (asa), a voiceless sibilant (aʒa), and a voiceless non-sibilant (afa) fricative, are illustrated in Figs. 3-5. The rate profile across AN fibers can be interpreted as a relatively straightforward distortion of the acoustic magnitude spectrum, with peaks and valleys in AN rates aligned with corresponding features in the spectrum, but with an overall rate profile that is shaped by the middle-ear filter to accentuate mid-frequency responses. However, the model IC response profiles differ markedly from the acoustic spectra. Model BE neurons have the highest rates near spectral slopes, where AN responses have relatively large fluctuations. Model BS neurons are excited by the channels with the least fluctuation, near the peaks in the spectra, but are suppressed by fluctuations associated with spectral slopes. Thus, the BS rate profile is “sharper” than the AN rate profile.

In the next section, the BE IC response panels will be used as a representation of the neural responses. In Figs. 1 and 2, BE IC responses encode voicing by a low-frequency peak below 250 Hz. Fricatives with the same place of articulation that differ only in voicing show similar BE IC response patterns that differ only in the low-frequency response. Sibilants show a double-peak response, while non-sibilants show a single-peak response at high frequencies. Postalveolar (/ʃ/, /ʒ/) sibilant’s first peak is broader and encompasses lower frequencies compared to alveolar sibilants (/s/, /z/). The sibilant /h/ shows a distinct absence of response beyond 8 kHz. Figure 1

shows that the unique double-peak vs. single pick response is mostly encoded by frequencies beyond 8 kHz (dashed line).

Figures 3 and 4 show that the processed conditions affect the voicing cue, and therefore, confusions across voiced vs. voiceless fricatives with the same place of articulation are to be expected. Furthermore, the differences in the high-frequency double-peak response between the sibilants with different places of articulation (alveolar vs. postalveolar) were diminished; the first peak was broader for the processed conditions. Thus, confusions within the sibilant category would be expected to increase for the processed conditions, particularly when the number of channels was reduced. Figure 5 shows that for non-sibilant fricatives, limiting the information to approximately 8 kHz in the processed conditions would introduce confusion with the fricative /h/.

Figure 1. Spectral and Neural Responses to Fricative-Sibilants

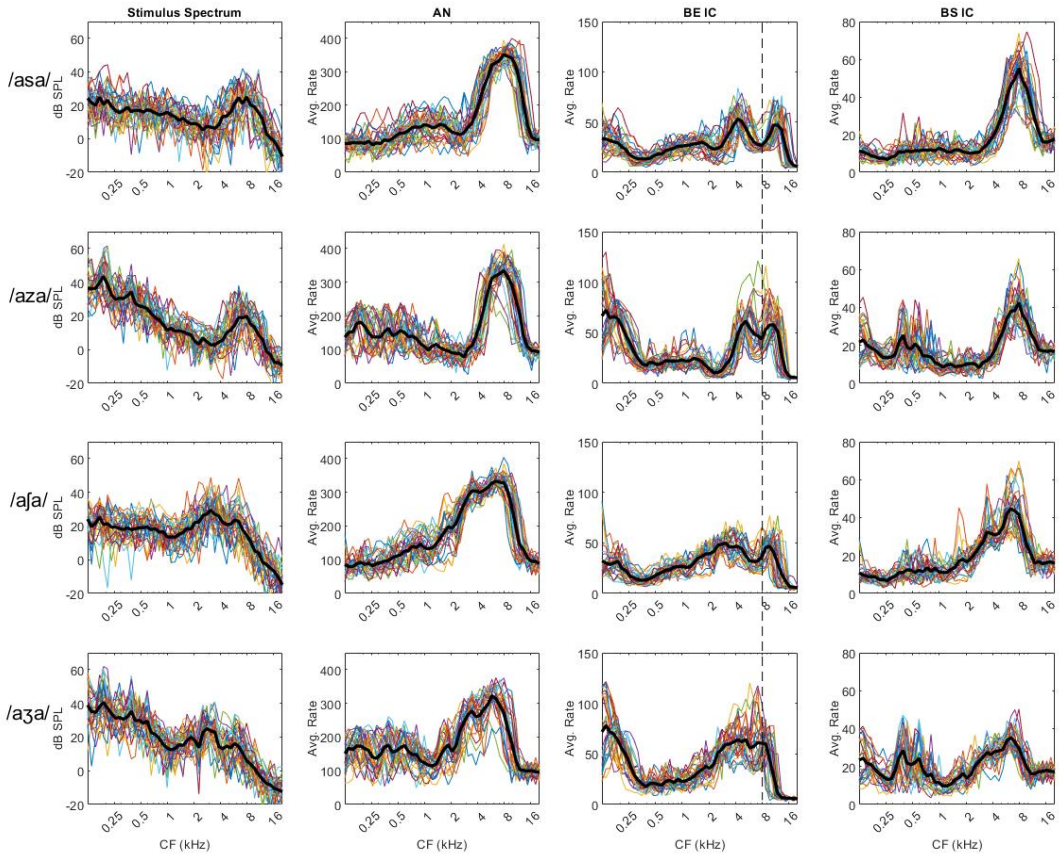

Figure 1 shows the mid-fricative spectrum and the corresponding neural response profiles for English sibilant fricatives. Panels from left to right demonstrate the spectra, auditory-nerve (AN), band-enhanced inferior colliculus (BE IC), and band-suppressed inferior colliculus (BS IC) responses. English sibilant fricatives from top to bottom are /s, /z/, /ʃ/, and/ʒ/. Each plot includes the 30 responses obtained from the 10 speakers. The black line indicates the average response.

Figure 2. Spectral and Neural Responses to Fricative-Non-Sibilants

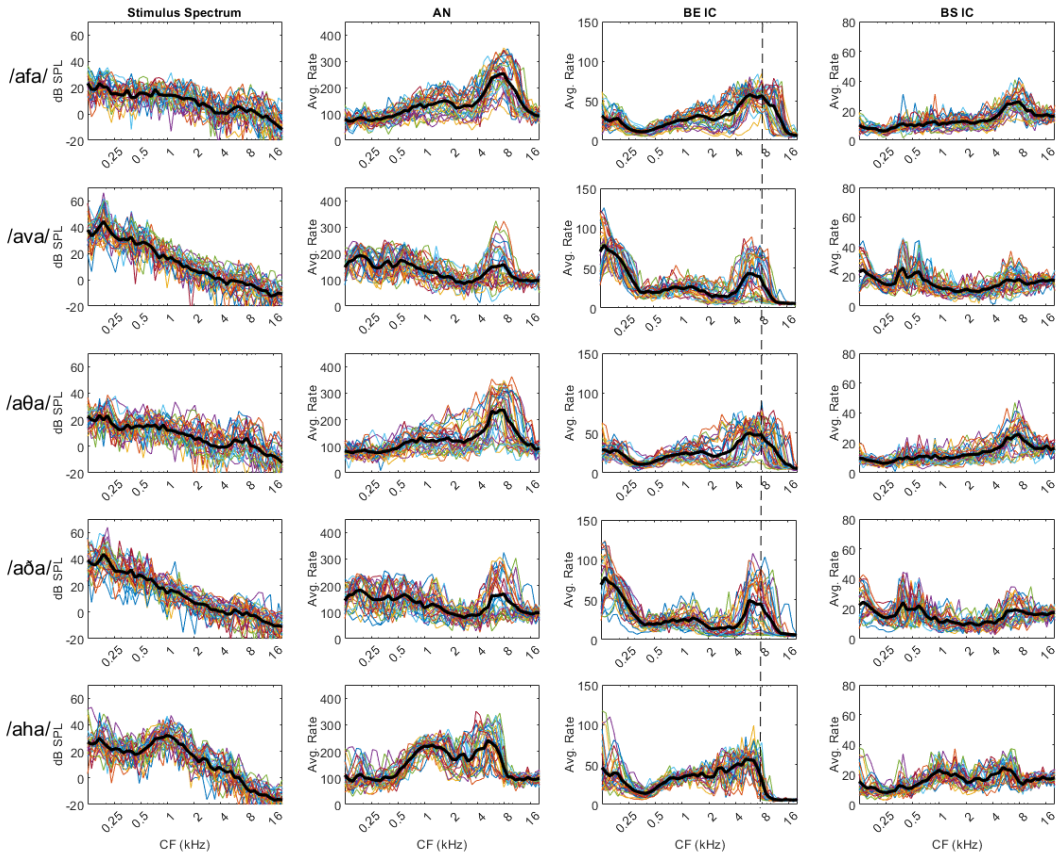

Figure 2 shows the mid-fricative spectrum and the corresponding neural response profile of English non-sibilant fricatives. Panels from left to right demonstrate the spectra, auditory-nerve (AN), band-enhanced inferior colliculus (BE IC), and band-suppressed inferior colliculus (BS IC) responses. English non-sibilant fricatives from top to bottom are /f/, /v/, /θ/, /ð/, and /h/. Each plot includes the 30 responses obtained from the 10 speakers. The black line indicates the average response.

Figure 3. Spectral and Neural Responses to /asa/ Across the Different Conditions

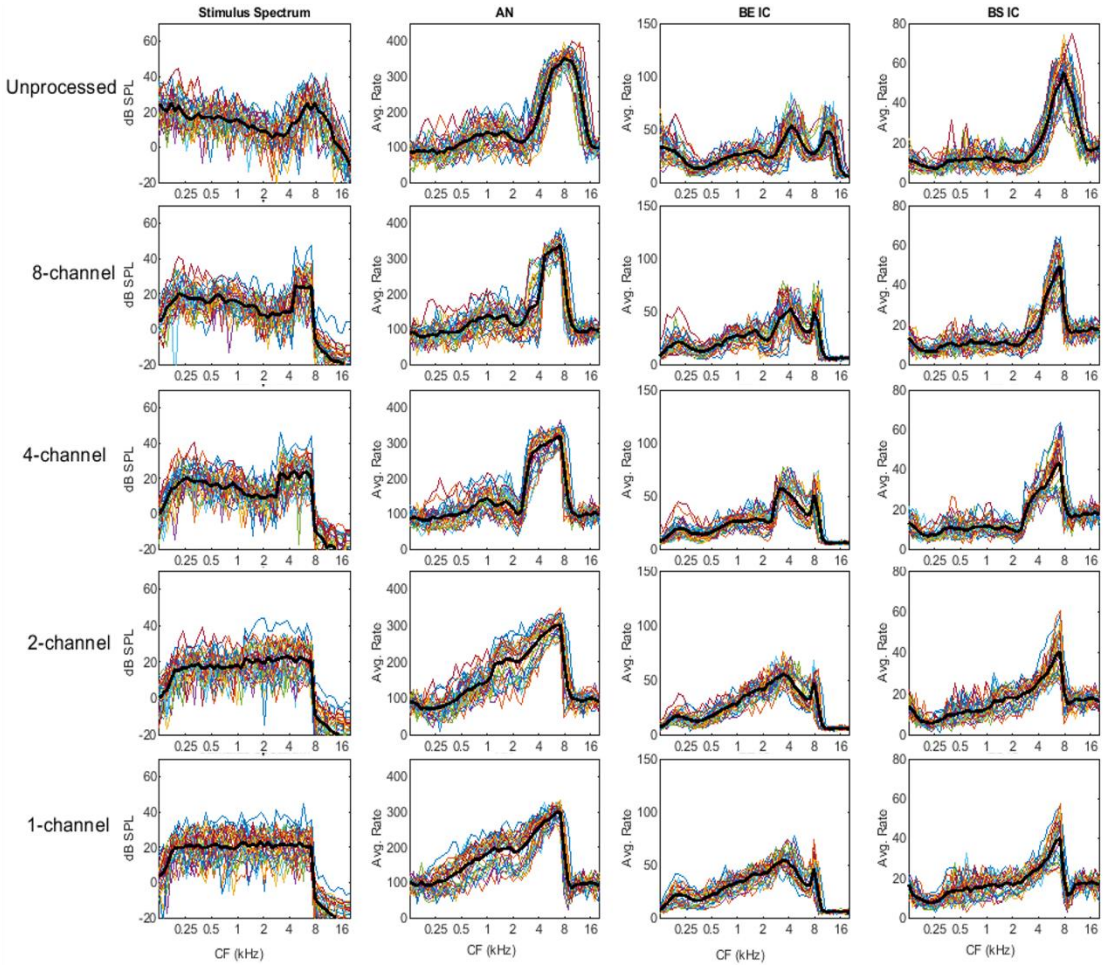

Figure 3 shows the mid-fricative spectrum and the corresponding neural response profile of sibilant fricative /asa/ across the different stimulus conditions. Panels from left to right demonstrate the spectra, auditory-nerve (AN), band-enhanced inferior colliculus (BE IC), and band-suppressed inferior colliculus (BS IC) responses. Stimulus conditions increase in spectral degradation from top to bottom, starting with the unprocessed stimuli, 8-channel, 4-channel, 2-channel, and 1-channel conditions. Each plot includes the 30 responses obtained from the 10 speakers. The black line indicates the average response.

Figure 4. Spectral and Neural Responses to /aʒa/ Across the Different Conditions

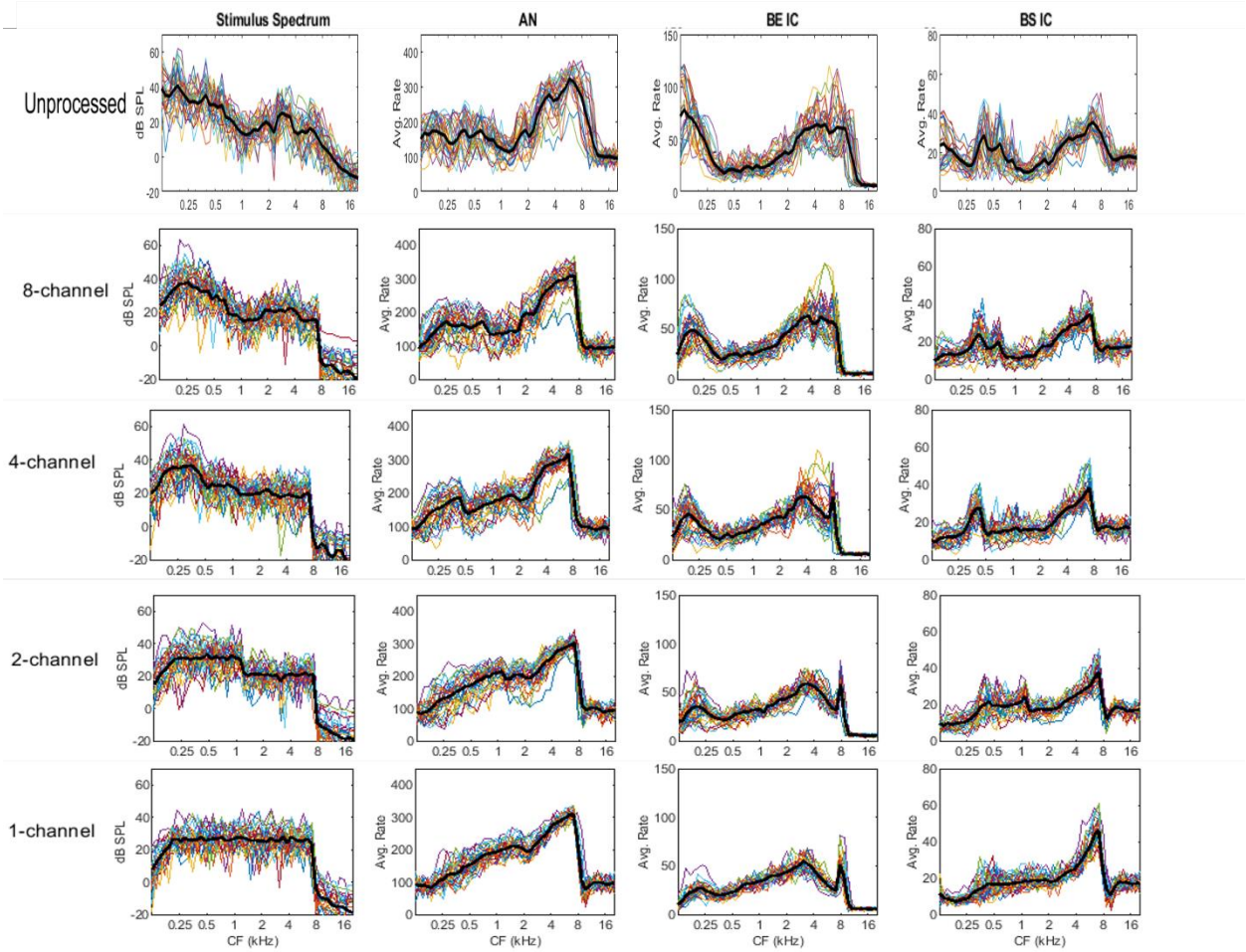

Figure 4 shows the mid-fricative spectrum and the corresponding neural response profile of sibilant fricative /aʒa/ across the different stimulus conditions. Panels from left to right demonstrate the spectra, auditory-nerve (AN), band-enhanced inferior colliculus (BE IC), and band-suppressed inferior colliculus (BS IC) responses. Stimulus conditions increase in spectral degradation from top to bottom starting with the unprocessed stimuli, 8-channel, 4-channel, 2-channel and 1-channel conditions. Each plot includes the 30 responses obtained from the 10 speakers. The black line indicates the average response.

Figure 5. Spectral and Neural Responses to /afa/ Across the Different Conditions

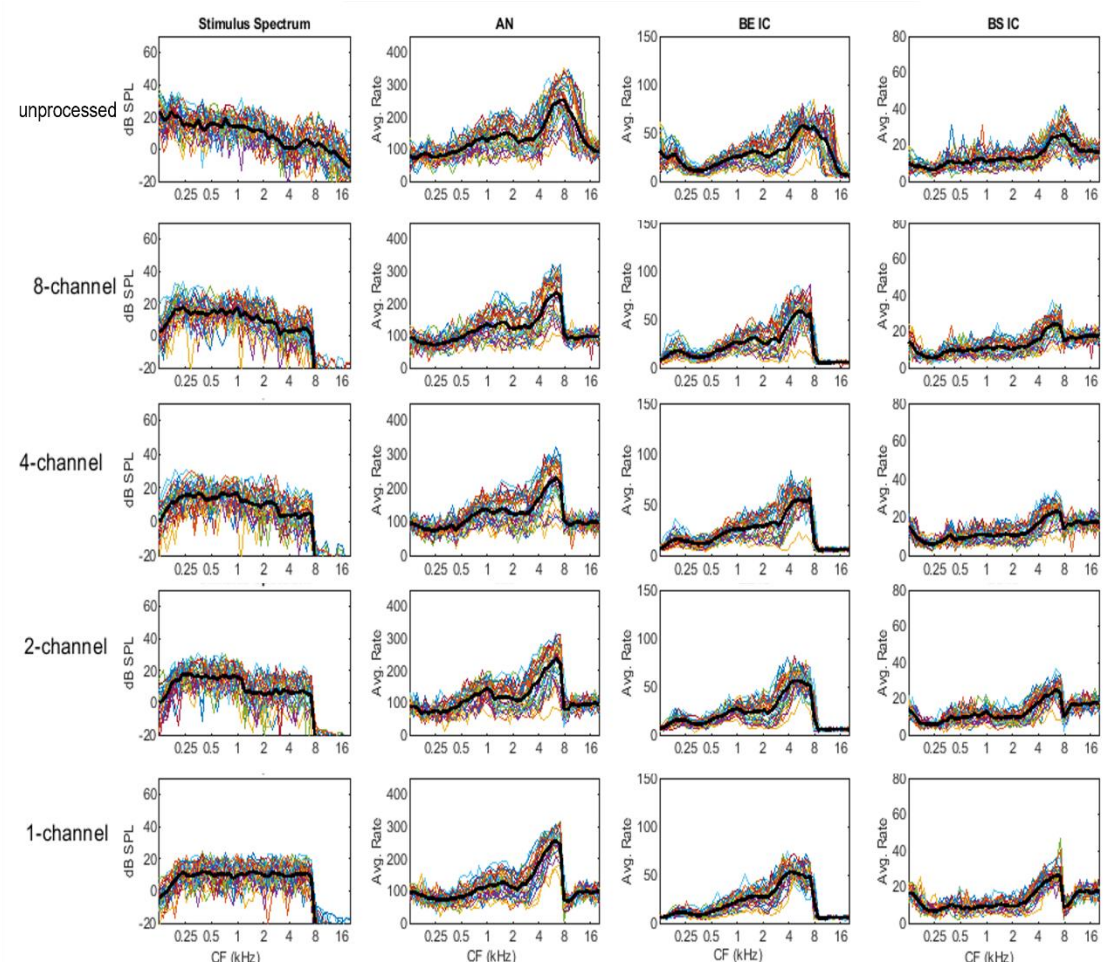

Figure 5 shows the mid-fricative spectrum and the corresponding neural response profile of non-sibilant fricative /afa/ across the different stimuli conditions. Panels from left to right demonstrate the spectra, auditory-nerve (AN), band-enhanced inferior colliculus (BE IC), and band-suppressed inferior colliculus (BS IC) responses. Stimulus conditions increase in spectral degradation from top to bottom starting with the unprocessed stimuli, 8-channel, 4-channel, 2-channel and 1-channel conditions. Each plot includes the 30 responses obtained from the 10 speakers. The black line indicates the average response.

A. Classifier-based analysis:

Tables I-VI demonstrate classifier-based and perceptual confusion matrices. Each table corresponds to one of the stimulus conditions. Tables I and II show the unprocessed conditions. Table I demonstrates the extended frequency (up to 20 kHz) and Table II the conventional (8 kHz) frequency condition. Tables III-VI show the processed conditions using 8, 4, 2, and 1 channels, respectively. Within each table, Panels A-D correspond to spectral, AN, BE, BS, BE+BS features. Panel E demonstrates the fricative perceptual confusion matrix derived from a larger confusion matrix including 16 VCV syllables. Scores are based on published results for a group of young listeners with normal hearing (Gallun and Souza, 2008). Bold-font accuracy scores indicate that the accuracy of identifying the fricative is above the mean classifier performance. Asterisks indicate the top four confusions within a specific confusion matrix. Table I shows that for the unprocessed extended frequency condition, the overall accuracy of the classifier-based analysis improved from stimulus spectrum (73.7%) to AN (83%) to BE (85.6%). On the other hand, for the unprocessed conventional frequency and processed conditions (Tables II-VI), the overall accuracy of the classifier-based analysis improved from stimulus to AN, but then fell at the level of the IC.

Compared to extended frequency data (Table I), limiting the data to 8 kHz had limited effect on the overall accuracy for the spectral (73.7% vs. 73.0%), AN (83.0% vs. 81.5%) and BS (79.6% vs. 79.3%) models. However, for the BE and the BE+BS IC models the effect was larger. The overall accuracies of the extended-frequency BE and BE+BS IC models were 85.6% and 84.1%, while the accuracies for the 8-kHz models were 72.6% and 77.4%, respectively. Apart from the differences in the overall accuracy between the two models, no noticeable differences were seen in the patterns of accuracies and confusions, i.e., in the 8-kHz model, fricatives were less accurately identified, but the pattern of accuracies was the same as for the-20 kHz model.

In the unprocessed condition, the highest accuracy was achieved by the BE features, whereas in the processed conditions, the BS or BE+BS features were more accurate than BE. Classifier based accuracies approached perceptual accuracy (90%) for the unprocessed, and 4- and 8-channel processed conditions. For the 2- and 1-channel processed conditions, model-based accuracies were slightly higher than the perceptual accuracies.

Looking closely at the fricative-specific accuracy scores in the unprocessed condition, neural responses provided robust correlates to perceptual accuracy. Sibilants were predicted with higher accuracies relative to non-sibilants (apart from /h/), similar to the perceptual accuracies, with the exception of /v/. However, in terms of confusions, the classifier-based analysis showed the highest confusions to be among the non-sibilant fricatives. Confusions were limited within the non-sibilant fricatives (/f/ with /θ/ and /v/ with /ð/), i.e., it was rare to confuse a non-sibilant with a sibilant. In terms of the perceptual data, the highest confusions were also among the non-sibilant fricatives; however, confusions of non-sibilant fricatives with sibilant fricatives were the highest (/θ/ with /s/ and /f/ with /ʃ/). The voicing cue appeared to be robust in the classifier-based analysis, with less confusions between voiced and voiceless fricatives. However, in perceptual data the voicing cue was not completely spared, as confusion of /ð/ with /θ/ was among the top 4 most common confusions.

In the processed conditions, sibilant accuracy appeared to be more affected than non-sibilant, as predicted from Fig. 3, where the differences in the high-frequency double-peak response between the sibilants with different places of articulation (alveolar vs. postalveolar) were diminished. This decline in sibilant accuracy increased for the degraded conditions, however, this trend was not in line with the perceptual accuracy for the less-degraded conditions (8 and 4-channels), where

sibilants were more accurately identified. These results may highlight the importance of the top-down processing in the difficult listening conditions. However, this top-down processing can improve the accuracy within limits (8 and 4 channels) but does not appear to play a significant role in the more degraded conditions (2 and 1 channel), where the classifier accuracy was in line with the perceptual accuracy. Additionally, as mentioned before, the BS model had the highest accuracy in the processed conditions. For confusions, as anticipated from Fig. 5 for non-sibilant fricatives, limiting the information to 8 kHz, to match the frequency range used in the perceptual task, resulted in confusions with the fricative /h/. The perceptual data lacked the fricative /h/.

Table I Classifier-based confusion matrices using the unprocessed stimuli.

Classifier-Based Prediction (%)									
	/s/	/z/	/ʃ/	/ʒ/	/f/	/v/	/θ/	/ð/	/h/[a]
stimuli	**A- Spectral: Overall Accuracy 73.7**								
/s/	**80**	3	3		7		7		
/z/	3	**97**							
/ʃ/			**100**						
/ʒ/				**93**		7			
/f/	3				63	3	30*		
/v/						27	3	70*	
/θ/			3		37*	3	53	3	
/ð/		3			3	40*		53	
/h/								3	**97**
B- AN: Overall Accuracy 83.0									
/s/	**100**								
/z/	3	**97**							
/ʃ/			**100**						
/ʒ/				**97**		3			
/f/					70		30*		
/v/		3				53		43*	
/θ/					17*		83		
/ð/		3			3	47*		47	
/h/									**100**
C- BE IC: Overall Accuracy 85.6									
/s/	**100**								
/z/		**100**							
/ʃ/			**100**						
/ʒ/			3	**93**		3			
/f/					77		23*		
/v/						63		37*	
/θ/					20*		80		
/ð/					3	37*	3	57	
/h/									**100**
D- BS IC: Overall Accuracy 79.6									
/s/	**97**		3						
/z/	3	**93**				3			
/ʃ/			**100**						
/ʒ/				**97**				3	
/f/					73		23*		3
/v/					3	53		43*	
/θ/			7		37*		57		
/ð/					3	50*		47	
/h/									**100**
E- BE+ BS IC: Overall Accuracy 84.1									
/s/	**97**		3						
/z/	3	**100**							
/ʃ/			**100**						
/ʒ/				**97**		3			
/f/					80		20*		
/v/						70		30*	
/θ/			3		33*		63		
/ð/		3			3	40*		53	
/h/					3				**97**
F- Perception: Overall Accuracy 90.8									
/s/	**99**							1	
/z/		**100**							
/ʃ/			**100**						
/ʒ/				**100**					
/f/	6		15*		79				
/v/						**98**		2	
/θ/	19*				8*		66	6	1
/ð/		3				5	7*	85	

Table II Classifier-based confusion matrices using the unprocessed stimuli with an 8 kHz cut-off.

Classifier-Based Prediction (%)									
	/s/	/z/	/ʃ/	/ʒ/	/f/	/v/	/θ/	/ð/	/h/[a]
stimuli	**A- Spectral: Overall Accuracy 73.0**								
/s/	**90**	3			3		3		
/z/		**97**		3					
/ʃ/			**97**		3				
/ʒ/		3	3	**87**				7	
/f/					57		43*		
/v/					3	27	3	67*	
/θ/	3				44*	3	47	3	
/ð/					3	30*	3	63	
/h/	3				3				**94**
B- AN: Overall Accuracy 81.5									
/s/	**93**	3					3		
/z/		**100**							
/ʃ/			**100**						
/ʒ/			3	**94**		3			
/f/	3				80		17*		
/v/		3			3	60		34*	
/θ/					43*		57		
/ð/						47*	3	50	
/h/									**100**
C- BE IC: Overall Accuracy 72.6									
/s/	**90**		4	3				3	
/z/		**97**				3			
/ʃ/	13		**87**						
/ʒ/				**94**	3	3			
/f/	3		3		64		30*		
/v/			7			43		50*	
/θ/	3		7		43*		47		
/ð/		3			3	44*		50	
/h/				3	13				**84**
D- BS IC: Overall Accuracy 79.3									
/s/	**93**				7				
/z/		**100**							
/ʃ/			**100**						
/ʒ/		4		**90**		3		3	
/f/					70		30*		3
/v/					3	57		40*	
/θ/			7		30*		63		
/ð/					3	57*		40	
/h/									**100**
E- BE+ BS IC: Overall Accuracy 77.4									
/s/	**87**		4	3	3		3		
/z/		**97**						3	
/ʃ/			**100**						
/ʒ/			3	**93**		3			
/f/	10				57		33*		
/v/		3				57		40*	
/θ/			7		23*		70		
/ð/		3			4	47*		47	
/h/					7	3			**90**
F- Perception: Overall Accuracy 90.8									
/s/	**99**							1	
/z/		**100**							
/ʃ/			**100**						
/ʒ/				**100**					
/f/	6		15*		79				
/v/						**98**		2	
/θ/	19*				8*		66	6	1
/ð/		3				5	7*	85	

Table III Classifier-based confusion matrices using the 8-channel stimuli.

Classifier-Based Prediction (%)									
	/s/	/z/	/ʃ/	/ʒ/	/f/	/v/	/θ/	/ð/	/h/[a]
stimuli	**A- Spectral: Overall Accuracy 50.3**								
/s/	30		33*	3	20				13
/z/	20	33		3		20	3	13	7
/ʃ/			**100**						
/ʒ/		3	17	**73**				3	3
/f/			7		23		3		67*
/v/						40	27	30	3
/θ/			3		33*	3	7		53*
/ð/					7	20	20	50	3
/h/			3						**97**
B- AN: Overall Accuracy 59.2									
/s/	7		17		53*		20		3
/z/	3	**73**	3	3		10	3	3	
/ʃ/			**100**						
/ʒ/			23	**70**	3				3
/f/					30		47*		23
/v/					0	**60**	33*	7	
/θ/			7		17		**63**		13
/ð/						37*	33*	30	
/h/									**100**
C- BE IC: Overall Accuracy 37.8									
/s/	7		53*		3		13		23
/z/	3	7	7		7	33*	27	7	10
/ʃ/			**90**						10
/ʒ/			13	20					67*
/f/			3		3		33*		60*
/v/						23	23	33*	20
/θ/			10				**57**		33
/ð/						20	27	33	17
/h/									**100**
D- BS IC: Overall Accuracy 55.2									
/s/	20		47*		23		10		
/z/	3	40	13		10	7	7	20	
/ʃ/			**100**						
/ʒ/		3	33*	**57**		3			3
/f/					**73**		20		7
/v/					17	37	27	20	
/θ/			3		43*		43		10
/ð/					20	17	33*	30	
/h/					3				**97**
E- BE+ BS IC: Overall Accuracy 41.4									
/s/	0		70*		3		13		13
/z/	7	10	7		10	33*	20	3	10
/ʃ/			**100**						
/ʒ/			20	40					40*
/f/			7		0		30		63*
/v/						30	20	33*	17
/θ/			13				**53**		33*
/ð/						23	23	40	13
/h/									**100**
F- Perception: Overall Accuracy 69.4									
/s/	**81**	3	3		7		6		
/z/	3	**70**		9		6		12	
/ʃ/	1		**99**						
/ʒ/				**100**					
/f/	10				69	1	20*		
/v/		6				59		23*	12
/θ/	19		1		46*		31	1	2
/ð/		6				27*	6	46	15

Table IV Classifier-based confusion matrices using the 4-channel stimuli.

Classifier-Based Prediction (%)									
	/s/	/z/	/ʃ/	/ʒ/	/f/	/v/	/θ/	/ð/	/h/[a]
stimuli	**A- Spectral: Overall Accuracy 36.1**								
/s/	3		77*		3				17
/z/	10	3	17	47*		3		10	10
/ʃ/			**93**						7
/ʒ/			27	**37**		3		3	30
/f/					23				77*
/v/				7		30	17	20	27
/θ/			7	3	37		13		40*
/ð/				7	7	27	20	23	17
/h/									**100**
B- AN: Overall Accuracy 47.3									
/s/	0		57*		30		10		3
/z/		10	13	50*	3			17	7
/ʃ/			**100**						
/ʒ/			33	43*	3				20
/f/					37		47*		17
/v/				3	3	**63**	23	7	
/θ/			7		30		**60**		3
/ð/				3	7	40	37	13	
/h/									**100**
C- BE IC: Overall Accuracy 29.2									
/s/	0		60*		3		7		30
/z/		0	33	3					63*
/ʃ/			**100**						
/ʒ/			27	7					67*
/f/			3		3		30		63*
/v/						0	27	27	47
/θ/			3				**57**		40
/ð/						3	33	13	50
/h/			17						**83**
D- BS IC: Overall Accuracy 32.6									
/s/	0		67*		27		7		
/z/		0	37	23	13		3	20	3
/ʃ/			**100**						
/ʒ/			37	**43**	7				13
/f/					**70**		27		3
/v/					17	**33**	2	27	3
/θ/			3		43*		**47**		7
/ð/					20			0	80*
/h/					17	27	17	40*	0
E- BE+ BS IC: Overall Accuracy 34.1									
/s/	10		70*				7		13
/z/		0	33	13			3		50*
/ʃ/			**100**						
/ʒ/			40	20					40
/f/			3		0		30		67*
/v/						20	23	17	40
/θ/			17				**50**		33
/ð/						10	37	10	43*
/h/			3						**97**
F- Perception: Overall Accuracy 46.6									
/s/	**53**	2	19		13	2	11		
/z/	4	**54**		22		15	1	4	
/ʃ/	49*		**49**		2				
/ʒ/	4	33*	2	**54**		5		2	
/f/	9				**63**		22	4	2
/v/	1	2				**64**		9	24
/θ/	14	1	2		49*	5	24	2	3
/ð/		2			1	65*	3	14	15

Table V Classifier-based confusion matrices using the 2-channel stimuli.

Classifier-Based Prediction (%)									
	/s/	**/z/**	**/ʃ/**	**/ʒ/**	**/f/**	**/v/**	**/θ/**	**/ð/**	**/h/[a]**
stimuli	**A- Spectral: Overall Accuracy 30.4**								
/s/	0		80*						20
/z/		0	17	7		13		3	60*
/ʃ/			**97**						3
/ʒ/			20	17		3			60*
/f/			3		30		3		63*
/v/						**33**	3	3	60*
/θ/			10		37		10		43
/ð/						2	13	7	60*
/h/			13	7					**80**
B- AN: Overall Accuracy 36.3									
/s/	0		53*		27		3		17
/z/		0	10	7	3				80*
/ʃ/			**87**						13
/ʒ/			7	10					83*
/f/					**40**		57*		3
/v/						30	23	3	43
/θ/			3		20		**67**		1
/ð/						17	33	3	47
/h/					10				**90**
C- BE IC: Overall Accuracy 26.2									
/s/	0		60*						40
/z/		0	33						67*
/ʃ/			**97**						3
/ʒ/			43	0					57
/f/			7		0		37		57
/v/						3	23		73*
/θ/			10				**60**		30
/ð/						7	27	3	63*
/h/			27						**73**
D- BS IC: Overall Accuracy 29.2									
/s/	0		63*		33				3
/z/		0	23		3	7			67*
/ʃ/			**93**						7
/ʒ/			10	10					80*
/f/					**37**		50		13
/v/					7	13	13	7	60*
/θ/			3		47		**43**		7
/ð/					13	20	13	0	53
/h/			3	3	20		7		**67**
E- BE+ BS IC: Overall Accuracy 25.7									
/s/	0		77*						23
/z/		0	37						63*
/ʃ/			**100**						
/ʒ/			33	3					63*
/f/			17		0		30		53
/v/						3	20		77*
/θ/			13				**50**		37
/ð/						7	20	0	73*
/h/			23						**77**
F- Perception: Overall Accuracy 21.3									
/s/	10		10		71*		9		
/z/		5	3	15	10	60*	2	5	
/ʃ/	19		**41**	1	35		4		
/ʒ/	5	13	2	**29**		43	1	7	
/f/	1	1			**65**	3	21	4	5
/v/		1				**58**		5	36
/θ/	8				74*	1	12	1	4
/ð/		1			1	70*		4	24

Table VI Classifier-based confusion matrices using the 1-channel stimuli.

Classifier-Based Prediction (%)									
	/s/	**/z/**	**/ʃ/**	**/ʒ/**	**/f/**	**/v/**	**/θ/**	**/ð/**	**/h/[a]**
stimuli	**A- Spectral: Overall Accuracy 15.6**								
/s/	0		80*						20
/z/		0	83*						17
/ʃ/			**87**						13
/ʒ/			87*	0					13
/f/			30		**33**				37
/v/			77		7	0			17
/θ/			37		40		0		23
/ð/			87*		3			0	10
/h/			80*						**20**
B- AN: Overall Accuracy 21.1									
/s/	0				33				67
/z/		0			17				83*
/ʃ/			0		0				100*
/ʒ/				0	0				100*
/f/					**43**		53		3
/v/					30	0			70*
/θ/					23		**67**		10
/ð/					40		3	0	57
/h/					20				**80**
C- BE IC: Overall Accuracy 17.9									
/s/	0		60						40
/z/		0	73*						27
/ʃ/			**97**						3
/ʒ/			97*	0					3
/f/			0		0		3		70
/v/			77*			0			23
/θ/			10				**43**		47
/ð/			57					0	43
/h/			80*						20
D- BS IC: Overall Accuracy 21.9									
/s/	0		40		33				27
/z/		0	67*		13				20
/ʃ/			**50**		3				47
/ʒ/			70*	0	3				27
/f/					**73**		23		3
/v/			57*		23	0	3		17
/θ/			7		40	3	**47**		3
/ð/			40		33		7	0	20
/h/			53*		17		3		**27**
E- BE+ BS IC: Overall Accuracy 15.2									
/s/	0		83*						17
/z/		0	97*						3
/ʃ/			**100**						
/ʒ/			93*	0					7
/f/			23		0		2		57
/v/			83*			0			17
/θ/			20				**27**		53
/ð/			60					0	40
/h/			90*						10
F- Perception: Overall Accuracy 18.3									
/s/	**34**	2			30		18	7	9
/z/	21	10	3	5	6	28	7	15	5
/ʃ/	35*	3	11		26	1	16	3	5
/ʒ/	18	9	2	5	4	40*	6	9	7
/f/	14	1	1		**19**	1	20	6	38*
/v/	9	11	1	4		**33**	6	15	21
/θ/	13		1	2	19	9	16	10	30
/ð/	5	19	1		2	41*	1	**19**	12

Sections A-F show the confusion matrices generated from spectral, AN, BE IC, BS IC, BE& BS IC features, and perceptual data (Gallun & Souza, 2008) in that order.

[a] perceptual data was derived from a larger confusion matrix including 16 VCV syllables. The fricative /h/ was not included in the 16 consonants. Confusions of fricatives with consonants of other manner was reported within /h/ column for perceptual data. Scores were obtained from a group of young listeners with normal hearing.

Bold cells indicate that the accuracy of identifying the fricative is above the mean classifier performance. * indicate the top four confusions within a specific confusion matrix.

IV. DISCUSSION:

Fricatives have been shown to exhibit considerable contextual and intra/inter-subject variability (Jongman et al., 2000; Narayanan et al., 1995; Shadle, 1990), as well as asymmetries in fricative voiced and voiceless contrast patterns cross linguistically (Chodroff and Wilson, 2020; Maddieson and Disner, 1984; Stevens, 1998). Yet, despite this variability and these patterns, fricative contrasts are perceived with high accuracy (Cutler et al., 2004; Gallun and Souza, 2008; Pisoni and Luce, 1986; Woods et al., 2010). In accounting for this variability, much, if not all, of this work has focused on the production system, the articulation and acoustic analysis of fricatives. Perception studies are behavioral studies, with the analyses based on the acoustic signal, the output of sound production and the input to the auditory system. This waveform is a mechanical signal transformed at the AN into tonotopic and non-linear signal that moves through the complex neural pathways of the auditory system. While the acoustic analyses are a powerful and productive tool, they lend little insight into the neural coding of speech, or the constraints imposed on perception by the auditory system. The inattention to auditory representation is linked to the lack of realistic models

of the neural responses to speech in the auditory pathways. Realistic neural models have been developed to model the coding of speech in the midbrain (Carney, 2018; Carney and McDonough, 2019; Nelson and Carney, 2004; Zilany et al., 2014; Zilany et al., 2009).

The current study used computational models for AN fibers and IC neurons to demonstrate the auditory representation of fricatives and establish the link between the production and perception of fricatives. The modeled neural responses, from 10 different speakers uttering each fricative 3 times, were compared to perceptual accuracy from a study by Gallun and Souza (2008). We hypothesized that the modeled neural responses can aid in developing explanations and representations of perceptual accuracies than the spectra of the acoustical signal alone.

A. Classifier vs. perceptual performance

The classifier analysis supported our hypothesis in the unprocessed condition (natural speech), for which the overall accuracy improved from the stimulus (73.7%) to AN (83%) to IC (BE, 85.6%) approaching perceptual accuracy (90%). The neural modeled accuracies remained below perceptual accuracies, as expected assuming the perceptual task benefits from higher-level neural equivalents of categorical perception (Lago et al., 2015), and top-down processing (Davis and Johnsrude, 2007). If anything, a larger gap between the classifier accuracy and perceptual accuracy was expected. The high accuracy in the classifier results could be explained by the smaller number of phoneme choices in the classifier (n=9) vs. perceptual (n=16) task.

For the processed conditions, the overall accuracy of the classifier analysis improved from the stimulus spectra to AN but then declined at the level of the IC, which could imply a greater role of top-down processing in degraded conditions. Top-down processing is thought to involve selective favoring of features based on prior knowledge. Highly-weighted features influence the

processing and encoding of sensory input (Asilador and Llano, 2021; Von Helmholtz, 1867). The neural modeled accuracies remained below perceptual accuracies for less degraded conditions (8- and 4-channel). However, for the more degraded conditions (2- and 1-channel), perceptual accuracies were slightly lower than the neural modeled accuracies. The higher accuracy in the classifier-based models in the more degraded conditions could be, once again, explained by the difference in task design (larger set in the perceptual task) and by the limits of top-down processing for strongly degraded stimuli.

B. Fricative-specific performance

The fricative-specific accuracy scores of the modeled neural responses were consistent with perceptual accuracy. Sibilants were predicted more accurately than non-sibilants. Classifier-based accuracy scores matched the behavioral data (apart from /v/). Other behavioral studies have shown controversial results as to whether /v/ is highly confusable (Cutler et al., 2004; Woods et al., 2010).

The unprocessed classifier confusions were not consistent with the perceptual data, suggesting that different processes may be involved in accuracy vs. confusions. While the accuracy of identification of a fricative may rely on encoding at the level of the IC, the pattern of fricative confusions might rely on other processes, including language phonotactics, experience, and top-down processing.

The results for different IC cell types in the classifier analysis varied with the acoustic signal quality; thus, the relative role of different IC cell types may vary in different listening environments. For the unprocessed condition, the BE features yielded the highest accuracy, whereas, in the processed condition, the BS or BE+BS IC features had the highest accuracy. The two groups of IC neurons represent stimulus features with opposite polarities, similar to the retinal

ganglion cells with on-center-off-surround and off-center-on-surround receptive fields (Carney and McDonough, 2019; Kim et al., 2020).

C. Hidden hearing loss

In the current study, we compared the classifier performance for the spectra or model responses limited to 8 kHz and 20 kHz to understand the role of extended frequencies on accuracy at different levels from stimulus to the IC. Results showed that extended frequencies only affected the accuracy performance at the level of the BE IC. It is generally assumed that frequencies limited to 8 kHz contain much of the phonetic cues, which has reduced the focus on the importance of the extended frequencies (Vitela et al., 2015). The results based on the stimulus spectra and AN models did not benefit from the additional information provided by the extended frequencies, whereas the BE IC results did. Limiting the frequency information did not seem to affect the fricative specific accuracies and confusions, i.e., it reduced the overall accuracy equally. The current study corroborated the growing evidence suggesting that acoustic information in the higher-frequency regions affects speech intelligibility (Badri et al., 2011; Levy et al., 2015; Moore et al., 2017; Pienkowski, 2017; Zadeh et al., 2019).

D. Conclusions, limitations, and future work

The current study showed that the modeled neural responses at the level of AN and IC provided better predictions of perceptual accuracies than did the stimulus spectra. Accuracies of fricative contrasts were explained by modeled neural response profiles, while confusions were only partially explained. Different IC cell types may play different roles in perception, based on the quality of the acoustic signals. Future studies measuring electrophysiological responses from different IC cell types may shed light on these roles. Extended frequencies (8-20 kHz) improved accuracies at

the level of IC only, explaining some of the discrepancy between acoustical and perceptual data in hidden hearing loss.

It is important to note that the modeled neural responses were limited to a single sound level. It is possible that by varying the sound level over a wide (realistic) range, the models could perform better, in comparison to the AN. However, perceptual data for comparison to model responses at different sound levels were not available. Other model limitations include the strictly on-CF implementation of the SFIE model and the lack of efferent control of cochlear gain, which might impact performance for degraded conditions (Farhadi et al., 2021). Comparison of the confusions in the processed conditions was limited by the frequency limit of 8 kHz used in the perceptual task, which resulted in higher confusions for the fricative /h/ (the perceptual data lacked the fricative /h/). Future studies that simulate neural responses of the hearing-impaired ear to fricatives could guide the development of processing strategies aimed at reaching the normal-hearing neural targets.

Acknowledgements: Supported by NIH-DC010813 (LHC).

REFERENCES:

Asilador, A., Llano, D. A. (2021). Top-Down Inference in the Auditory System: Potential Roles for Corticofugal Projections. *Frontiers in Neural Circuits, 14*. doi: 10.3389/fncir.2020.615259

Badri, R., Siegel, J. H., Wright, B. A. (2011). Auditory filter shapes and high-frequency hearing in adults who have impaired speech in noise performance despite clinically normal audiograms. *The Journal of the Acoustical Society of America, 129*(2), 852-863.

Behrens, S., Blumstein, S. E. (1988). On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants. *The Journal of the Acoustical Society of America, 84*(3), 861-867.

Boersma, P., Weenink, D. (2016). Praat (Version 6.0. 23). *Amsterdam, document is copyrighted*.

Carney, L. H. (2018). Supra-threshold hearing and fluctuation profiles: implications for sensorineural and hidden hearing loss. *Journal of the Association for Research in Otolaryngology, 19*(4), 331-352.

Carney, L. H., Li, T., McDonough, J. M. (2015). Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations. *Eneuro, 2*(4).

Carney, L. H., McDonough, J. M. (2019). Nonlinear auditory models yield new insights into representations of vowels. *Attention, Perception, & Psychophysics, 81*(4), 1034-1046.

Catford, J. C. (1977). *Fundamental problems in phonetics*: Midland Books.

Chodroff, E., Wilson, C. (2020). Acoustic–phonetic and auditory mechanisms of adaptation in the perception of sibilant fricatives. *Attention, Perception, & Psychophysics, 82*(4), 2027-2048.

Chodroff, E., Wilson, C. (2022). Uniformity in phonetic realization: Evidence from sibilant place of articulation in American English. *Language*.

Crystal, T. H., House, A. S. (1988). Segmental durations in connected-speech signals: Syllabic stress. *Journal of the Acoustical Society of America, 83*, 1574-1585.

Cutler, A., Weber, A., Smits, R., Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America, 116*(6), 3668-3678.

Davis, M. H., Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research, 229*(1), 132-147. doi: https://doi.org/10.1016/j.heares.2007.01.014

Deng, L., Geisler, C. D. (1987). Responses of auditory-nerve fibers to nasal consonant–vowel syllables. *The Journal of the Acoustical Society of America, 82*(6), 1977-1988.

Evers, V., Reetz, H., Lahiri, A. (1998). Crosslinguistic acoustic categorization of sibilants independent of phonological status. *Journal of phonetics, 26*(4), 345-370.

Farhadi, A., Jennings, S. G., Strickland, E. A., Carney, L. H. (2021). *A Closed-Loop Gain-Control feedback model for the medial efferent system of the descending auditory pathway.* Paper presented at the ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Forrest, K., Weismer, G., Milenkovic, P., Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *The Journal of the Acoustical Society of America, 84*(1), 115-123.

Gallun, F., Souza, P. (2008). Exploring the role of the modulation spectrum in phoneme recognition. *Ear and hearing, 29*(5), 800.

Gordon, M., Barthmaier, P., Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association, 32*(2), 141-174.

Goutman, J. D., Glowatzki, E. (2007). Time course and calcium dependence of transmitter release at a single ribbon synapse. *Proceedings of the National Academy of Sciences, 104*(41), 16341-16346. doi: doi:10.1073/pnas.0705756104

Haggard, M. (1978). The devoicing of voiced fricatives. *Journal of phonetics, 6*(2), 95-102.

Hughes, G. W., Halle, M. (1956). Spectral properties of fricative consonants. *The journal of the acoustical society of America, 28*(2), 303-310.

Ibrahim, R. A., Bruce, I. C. (2010). Effects of peripheral tuning on the auditory nerve's representation of speech envelope and temporal fine structure cues *The neurophysiological bases of auditory perception* (pp. 429-438): Springer.

Jassem, W. (1979). Classification of fricative spectra using statistical discriminant functions. *Frontiers of speech communication research*, 77-91.

Jesus, L. M. T., Shadle, C. H. (2002). A parametric study of the spectral characteristics of European Portuguese fricatives. *J. Phonetics, 30*, 437-464.

Jongman, A., Wayland, R., Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America, 108*(3), 1252-1263.

Joris, P., Schreiner, C., Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological reviews, 84*(2), 541-577.

Kiang, N., Moxon, E. (1974). Tails of tuning curves of auditory-nerve fibers. *The Journal of the Acoustical Society of America, 55*(3), 620-630.

Kim, D. O., Carney, L., Kuwada, S. (2020). Amplitude modulation transfer functions reveal opposing populations within both the inferior colliculus and medial geniculate body. *Journal of Neurophysiology, 124*(4), 1198-1215.

Kim, D. O., Zahorik, P., Carney, L. H., Bishop, B. B., Kuwada, S. (2015). Auditory distance coding in rabbit midbrain neurons and human perception: monaural amplitude modulation depth as a cue. *Journal of Neuroscience, 35*(13), 5360-5372.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *The Journal of the Acoustical Society of America, 59 5*, 1208-1221.

Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection.* Paper presented at the Ijcai.

Krishna, B. S., Semple, M. N. (2000). Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *Journal of neurophysiology, 84*(1), 255-273.

Ladefoged, P. (1971). Preliminaries to Linguistic Phonetics. Chicago: University of Chicago Press. *Journal of Phonetics, 1*(1), 93-99.

Ladefoged, P., Maddieson, I. (1996). Sounds of the world's languages. Hoboken: NJ: Wiley-Blackwell.

Lago, S., Scharinger, M., Kronrod, Y., Idsardi, W. J. (2015). Categorical effects in fricative perception are reflected in cortical source information. *Brain and language, 143*, 52-58. doi: 10.1016/j.bandl.2015.02.003

Langner, G. (1992). Periodicity coding in the auditory system. *Hearing research, 60*(2), 115-142.

Levy, S. C., Freed, D. J., Nilsson, M., Moore, B. C., Puria, S. (2015). Extended high-frequency bandwidth improves reception of speech in spatially separated masking speech. *Ear and hearing, 36*(5), e214.

Liberman, M. C., Epstein, M. J., Cleveland, S. S., Wang, H., Maison, S. F. (2016). Toward a differential diagnosis of hidden hearing loss in humans. *PloS one, 11*(9), e0162726.

Maddieson, I., Disner, S. F. (1984). *Patterns of sounds*: Cambridge university press.

Miller, R. L., Schilling, J. R., Franck, K. R., Young, E. D. (1997). Effects of acoustic trauma on the representation of the vowel/ε/in cat auditory nerve fibers. *The Journal of the Acoustical Society of America, 101*(6), 3602-3616.

Moore, D., Hunter, L., Munro, K. (2017). Benefits of extended high-frequency audiometry for everyone. *The Hearing Journal, 70*(3), 50-52.

Moser, T., Beutner, D. (2000). Kinetics of exocytosis and endocytosis at the cochlear inner hair cell afferent synapse of the mouse. *Proceedings of the National Academy of Sciences, 97*(2), 883-888. doi: doi:10.1073/pnas.97.2.883

Narayanan, S. S., Alwan, A. A., Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *The Journal of the Acoustical Society of America, 98*(3), 1325-1347.

Nelson, P. C., Carney, L. H. (2004). A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. *The Journal of the Acoustical Society of America, 116*(4), 2173-2186.

Nelson, P. C., Carney, L. H. (2007). Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus. *Journal of neurophysiology, 97*(1), 522-539.

Pienkowski, M. (2017). On the etiology of listening difficulties in noise despite clinically normal audiograms. *Ear and Hearing, 38*(2), 135.

Pisoni, D. B., Luce, P. A. (1986). Speech perception: Research, theory and the principal issues. *Pattern recognition by humans and machines: Speech perception, 1*, 1-50.

Proctor, M. I., Shadle, C. H., Iskarous, K. (2010). Pharyngeal articulation in the production of voiced and voiceless fricatives. *The Journal of the Acoustical Society of America, 127*(3), 1507-1518.

Raman, I. M., Zhang, S., Trussell, L. O. (1994). Pathway-specific variants of AMPA receptors and their contribution to neuronal signaling. *Journal of Neuroscience, 14*(8), 4998-5010.

Rhode, W. S. (1971). Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique. *The Journal of the Acoustical Society of America, 49*(4B), 1218-1231.

Sachs, M. B., Abbas, P. J. (1974). Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli. *The Journal of the Acoustical Society of America, 56*(6), 1835-1847.

Schreiner, C. E., Winer, J. A. (2005). *The inferior colliculus*: Springer.

Schroeder, M. R. (1968). Period histogram and product spectrum: New methods for fundamental-frequency measurement. *The Journal of the Acoustical Society of America, 43*(4), 829-834.

Shadle, C. (1985). *The Acoustics of Fricative Consonants, unpublished Ph. D.* dissertation, MIT.

Shadle, C. H. (1990). Articulatory-acoustic relationships in fricative consonants *Speech production and speech modelling* (pp. 187-209): Springer.

Shadle, C. H., Badin, P., Mouliner, A. (1992). *Towards the spectral characteristics of fricative consonants.* Paper presented at the Twelfth International Congress of Phonetic Sciences.

Shera, C. A., Guinan, J. J., Oxenham, A. J. (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proceedings of the National Academy of Sciences, 99*(5), 3318-3323.

Silbert, N., De Jong, K. (2008). Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. *The Journal of the Acoustical Society of America, 123*(5), 2769-2779.

Stevens, K. N. (1998). *Acoustic Phonetics (Current studies in linguistics; 30)*: MIT press.

Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *The Journal of the Acoustical Society of America, 91*(5), 2979-3000.

Strevens, P. (1960). Spectra of fricative noise in human speech. *Language and speech, 3*(1), 32-49.

Vitela, A. D., Monson, B. B., Lotto, A. J. (2015). Phoneme categorization relying solely on high-frequency energy. *The Journal of the Acoustical Society of America, 137*(1), EL65-EL70.

Von Helmholtz, H. (1867). *Handbuch der physiologischen Optik: mit 213 in den Text eingedruckten Holzschnitten und 11 Tafeln* (Vol. 9): Voss.

Westerman, L. A., Smith, R. L. (1984). Rapid and short-term adaptation in auditory nerve responses. *Hearing research, 15*(3), 249-260.

WHO. (2008). Grades of hearing impairment.

Woods, D. L., Yund, E. W., Herron, T. J., Cruadhlaoich, M. A. U. (2010). Consonant identification in consonant-vowel-consonant syllables in speech-spectrum noise. *The Journal of the Acoustical Society of America, 127*(3), 1609-1623.

Yates, G. K. (1990). Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions. *Hearing research, 50*(1-2), 145-162.

Yates, G. K., Winter, I. M., Robertson, D. (1990). Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range. *Hearing research, 45*(3), 203-219.

Young, E. D., Sachs, M. B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *The Journal of the Acoustical Society of America, 66*(5), 1381-1403.

Zadeh, L. M., Silbert, N. H., Sternasty, K., Swanepoel, D. W., Hunter, L. L., Moore, D. R. (2019). Extended high-frequency hearing enhances speech perception in noise. *Proceedings of the National Academy of Sciences, 116*(47), 23753-23759.

Zilany, M. S., Bruce, I. C., Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *The Journal of the Acoustical Society of America, 135*(1), 283-286.

Zilany, M. S., 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. *The Journal of the Acoustical Society of America, 126*(5), 2390-2412.