Fluctuation contrast and speech-on-speech masking: Model midbrain responses to simultaneous speech

LAUREL H. CARNEY[1,2,*]

[1] *Departments of Biomedical Engineering, Neuroscience, and Electrical & Computer Engineering, Del Monte Institute for Neuroscience, University of Rochester, Rochester, NY, USA*

[2] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

At the level of the auditory midbrain, low-frequency fluctuations within each frequency channel drive neurons with band-pass modulation transfer functions (MTFs). The amplitude of low-frequency fluctuations in ascending neural signals is affected by stimulus amplitude due to the gradual saturation of the inner hair cells (IHCs) beginning at moderate sound levels. This level dependence of low-frequency fluctuation amplitudes results in contrast cues at the level of the midbrain: Spectral peaks result in lower responses of cells with bandpass-MTFs, whereas spectral valleys result in higher responses. Here, we focus on model population midbrain responses with different best-modulation frequencies (BMFs) to simultaneous speech. Midbrain responses were simulated for single hearing-in-noise (HINT) sentences and for a pair of simultaneous sentences, spoken by a male and a female. Correlations between population responses to individual male (or female) sentences and responses to simultaneous sentences vary with BMF in the range of the male (or female) fundamental frequencies. The pattern of fluctuation contrast across frequency in the midbrain representation provides a framework for studying speech-on-speech masking for listeners with normal hearing and sensorineural hearing loss.

INTRODUCTION

Neural responses to complex sounds such as speech convey a number of interacting cues to the central nervous system. A better understanding of which cues are most significant for encoding speech information should improve the ability to restore or enhance these cues for listeners with hearing loss. Furthermore, it is important to understand how candidate cues are affected by hearing loss as well as by typical challenges such as masking. In this study, we use computational models to explore fluctuation contrast cues in response to spoken sentences. In particular, we show examples of how these cues are affected by a competing voice, in anticipation of future tests that will take advantage of listener performance data collected using the Competing Voices Test (CVT; Bramsløw *et al.*, 2014) with the Danish Hearing-in-Noise (HINT) sentences (Nielsen and Dau, 2009). We also explore examples of model

*Corresponding author: laurel.carney@rochester.edu

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
 ISBN: 978-87-990013-6-1.

responses that illustrate how sensorineural hearing loss affects fluctuation contrast cues in response to single speakers and competing voices.

One goal of this work is to test the hypothesis that across-frequency contrasts in fluctuation provide cues for the locations of formants (Carney *et al.*, 2015) and other spectral features in speech. Fluctuation contrasts are set up in the auditory periphery, and ultimately provide a robust representation in the responses of midbrain neurons. Neurons in the auditory midbrain (inferior colliculus, IC) are sensitive to low-frequency fluctuations, or periodicities, in their inputs. Auditory-nerve (AN) fibres phase-lock to the low-frequency fine structure in stimuli, and they simultaneously phase-lock to low-frequency fluctuations in response to complex sounds (Joris and Yin, 1992; review: Joris *et al.*, 2004). The low-frequency fluctuations in AN responses to speech include both the fundamental frequency of voiced sounds and the features of cochlear-filter induced envelopes in response to noisy sounds (Joris, 2003) such as fricative consonants.

Low-frequency fluctuations in peripheral responses are not limited to low characteristic frequencies (CFs) but occur across the entire AN population. These fluctuations in the neural responses are conveyed via the brainstem to the midbrain. AN fibres that convey fluctuations may have saturated average discharge rates, especially in the case of the sensitive high-spontaneous-rate (HSR) AN fibres, but their temporal patterns still convey substantial information in the form of phase-locking to fine-structure and low-frequency fluctuations. These low-frequency fluctuations are effective in exciting (or suppressing) midbrain neurons (Krishna and Semple, 2000; Nelson and Carney, 2007; Kim *et al.*, 2015). In the IC, average discharge rates depend on the amplitudes of low-frequency fluctuations on their inputs. Thus, changes across frequency in the amplitude of peripheral fluctuations carried by AN fibres result in a profile of midbrain rates that vary across frequency.

The above description is focused on how AN fibres carry the fluctuations in complex sounds in to the central nervous system (CNS). However, in the healthy ear, the representation of spectral features by these fluctuations is strongly shaped by two nonlinearities in the inner ear. First, near spectral peaks, such as formants in voiced sounds, the saturation of inner hair cell (IHC) transduction results in a "flattening" of the low-frequency fluctuations, or envelope-related features. As a result, AN fibres near spectral peaks have relatively low-amplitude low-frequency fluctuations, and instead have temporal responses that are dominated by a single harmonic closest to the spectral peak. This phenomenon is referred to as "synchrony capture" because it has been quantified on the basis of fine-structure phase-locking (Miller *et al.*, 1997), but it could equally well be referred to as a suppression of the low-frequency fluctuation. Importantly, the saturation of IHCs in frequency channels near spectral peaks results in changes in fluctuation amplitude across frequency channels, referred to here as fluctuation contrasts.

The second inner-ear nonlinearity that plays a role in shaping fluctuation contrasts is compressive cochlear amplification, which determines the sensitivity of the organ of corti response, and thus the set-point of the IHC transduction nonlinearity. Because

(in the healthy ear) cochlear sensitivity is controlled in a frequency-dependent manner, fluctuation contrasts can occur at each spectral peak in a complex sound. For example, synchrony capture occurs for multiple formant peaks in AN vowel responses, not only at the highest magnitude peak (Delgutte and Kiang, 1984).

Contrasts in the amplitude of fluctuations in AN responses across frequency channels provide a code for spectral peak frequencies (Carney *et al.*, 2015), but this code requires both frequency-dependent cochlear amplification, driven by the outer hair cells (OHCs), and sensitive transduction by IHCs. Therefore, sensorineural hearing loss involving reduced sensitivity of OHCs and/or IHCs will result in decreased fluctuation contrasts. The impact of modelled sensorineural hearing loss on model IC population responses is explored here.

Previous studies of the fluctuation contrast cues have reported model IC and physiological responses for gaussian-noise maskers in normal-hearing rabbits (Carney *et al.*, 2015). Responses of models with sensorineural hearing loss to speech with additive Gaussian noise have also been described (Carney *et al.*, 2016). Here we extend this exploration to competing-voice maskers. We hypothesize that the ability to segregate speakers with different fundamental frequencies requires a) valid fluctuation contrast cues set up in peripheral responses, and b) midbrain neurons tuned to modulation frequencies in the range of voice pitch (e.g., Langner and Schreiner, 1988). Using a simple model for band-enhanced modulation tuning in the IC, we can study the ability to separate responses of speakers with different F0s based on midbrain responses.

MODEL METHODS

The example responses illustrated here were created using the Zilany *et al.* (2014) AN model as the input to a simple modulation filter model for IC neurons (Mao *et al.*, 2013). This IC model is a simplification of the same-frequency inhibition-excitation model of Nelson and Carney (2004); The modulation filter model allows more flexible selection of the IC best modulation frequency (BMF). Modulation filter responses were passed through a first-order low-pass filter (Fc = 500 Hz) to approximate the frequency limit of temporal following in the IC (Joris *et al.*, 2004).

Model responses with sensorineural hearing loss shown here were simulated by reducing the sensitivity of the AN model IHC by setting the parameter C_{IHC} to 0.2 for all CFs, and by reducing the cochlear amplification of the IHCs by setting the parameter C_{IHC} to 0.2 for all CFs. This simple strategy for simulating sensorineural hearing loss results in a model with sloping hearing loss that ranges from ~15-20 dB at low frequencies up to ~40 dB at high frequencies. More detailed audiometric configurations for comparison to individual listener results can be modelled by fine tuning these parameters as a function of frequency.

RESULTS

Responses of model populations of IC neurons driven by fluctuations in the normal-hearing AN model are shown in Fig. 1. Four population responses to HINT sentences

(Nilsson *et al.*, 1994) are illustrated. Figure 1A shows the response of a population of model IC cells to a HINT sentence spoken by a male. The IC model neurons have CFs from 200-6000 Hz, and all cells have BMF=100 Hz, in the range of the F0 for a male speaker. IC modulation filters are broad (quality index, Q=1); therefore, small changes in F0 over the course of a sentence do not cause large changes in the responses of these filters. This population of IC cells are most strongly driven by features associated with a male voice.

Figure 1B shows responses of IC model cells with BMF = 200 Hz, in the range of female voice pitch, in response to a HINT sentence spoken by a female. In Figs. 1A-1B, the formant frequencies during voiced portions of the sentences are indicated by white circles, which represent frequencies at which the responses to fluctuations are suppressed due to IHC saturation. The yellow (white) regions represent frequency channels that respond strongly to low-frequency fluctuations that are set up in the periphery. Note that the AN-fibre average rates are saturated across all of the low to mid frequencies during the voiced sounds, but the differences in the amplitudes of the low-frequency fluctuations result in strong formant-related features in the model IC responses. During the consonants (e.g., cyan square), fluctuations are strongest in frequency bands associated with the rising slope of the stimulus spectra, rather than at the peak frequencies where IHCs are often saturated and thus fluctuation amplitudes are reduced.

Figures 1C and 1D show responses of the same IC population models to simultaneous presentation of the two sentences from Figs. 1A and 1B. In both panels, fluctuation contrast features from both sentences are apparent. However, the IC population tuned to BMFs in the range of male pitch (Fig. 1C) is dominated by features in the sentence spoken by the male (e.g., white circles). Likewise, in the model cells tuned to an F0 in the female range (Fig. 1D), the fluctuation contrast features are most similar to those in the response to the female sentence (e.g., orange circles).

To quantify the degree of similarity of the response to the masked sentence to the response to single sentences, correlations between the two-dimensional (CF × time) images were computed. Similar to conventional studies of masking, a comparison between the response to the target plus masker and the response to the masker alone provides an indication of how well the target can be segregated from the masker based on this representation. In this case, a lower correlation between two images indicates better separability of the target from the masker based on the fluctuation contrast features in the IC responses.

Figure 2 shows the correlations between T+M and M-alone as a function of the BMF of the IC neurons. Each point in Fig. 2A represents the correlation of two 2D images of IC population responses, such as those shown in Fig. 1. IC cells with BMFs in the range of the male voice have the largest differences (lowest correlations) between the response to the male speaker alone and the response to the simultaneous sentences. Similarly, the responses of IC cells with BMFs near the female F0 would be best able to segregate the female-alone response from the competing-voice masker.

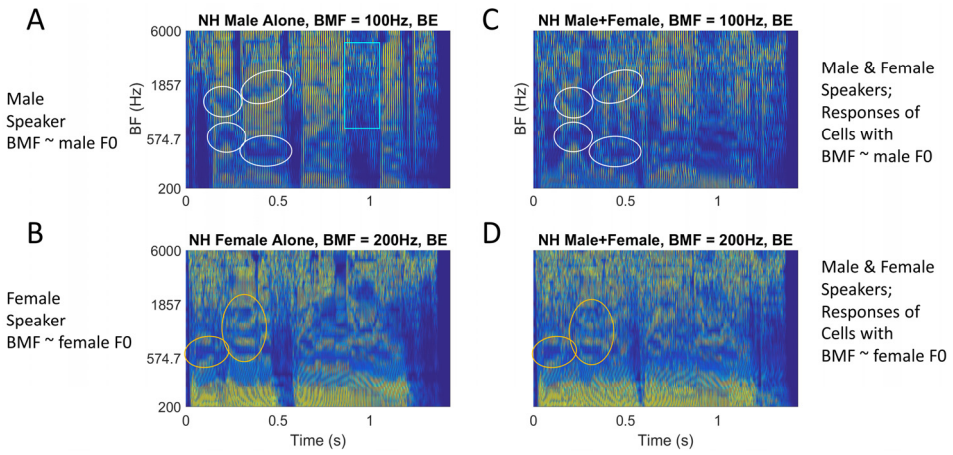

Fig. 1: Examples of population IC model responses to HINT sentences. Model responses here received inputs a population of normal-hearing (NH) high-spontaneous-rate AN fibres with characteristic frequencies ranging from 200-6 kHz. Each IC model is a bandpass modulation filter; best modulation frequencies are in the range of F0s for male (100 Hz, A, C) or female (200 Hz, B, D) speakers. A) Responses of a 100-Hz BMF IC model population to a male saying "Strawberry jam is sweet." White circles highlight response features for F1 and F2 in the first 2 words. The cyan square highlights the response to a fricative. B) Responses of a 200-Hz BMF IC population to a female of saying "They heard a funny noise." Orange circles highlight a few response features. C) 100-Hz BMF population response to simultaneous presentation of the sentences in A and B. D) 200-Hz BMF population response to the simultaneous sentences from A and B. Qualitatively, the response to the combined sentences contain fluctuation features from both sentences, but the neurons with 100-Hz BMF represent features that are more similar to the male-alone responses (e.g., white circles), and the 200-Hz responses are more similar to the female-alone response (e.g., orange circles). Sentences are from the English HINT test (Nilsson *et al.*, 1994). (color online)

Figure 2B shows similar calculations for IC models that include sensorineural hearing loss (SNHL) in the model AN inputs. At first glance, the more linear models for SNHL ear result in responses to the voices that appear to be more separable; However, the representation of many features in the sentences are reduced by decreased sensitivity. Figure 3 shows population responses for the models with SNHL; These responses include fewer cross-frequency contrasts associated with vowel formants, and very little response to consonants, as compared to responses of the normal-hearing model (cf. Fig. 1). The dotted lines in Fig. 2B show the same calculations after linear amplification of the speech inputs by 20 dB. This amplification restores the representation of some features in the populations responses (not shown), but decreases the separation of the correlations shown in Fig. 2B.

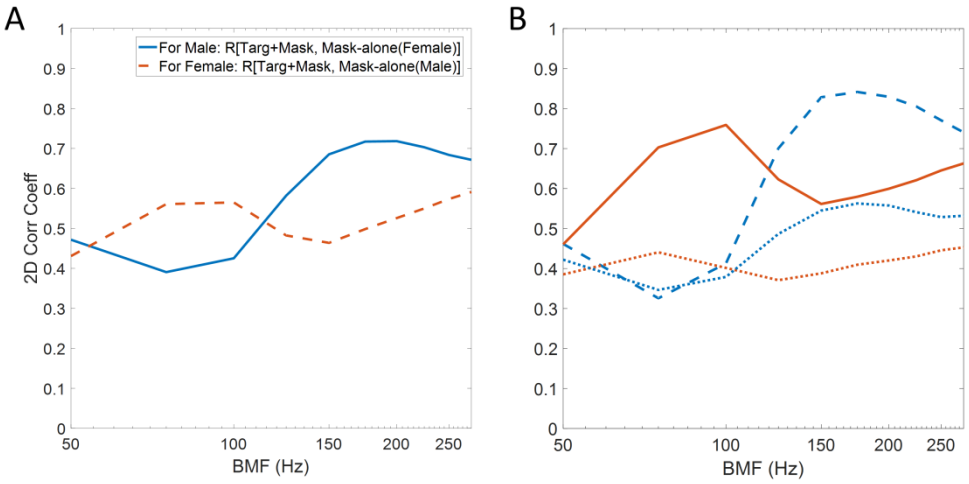

Fig. 2: Correlations between IC responses at each BMF to Target + Masker (i.e. two simultaneous sentences) and Masker alone (one sentence). Sentences are the same as those illustrated in Fig. 1. A) Blue: For the segregation of the male voice, the correlation considered is that between simultaneous sentence (Target + Masker) and Female (Masker alone). This correlation is lowest for IC population responses that have BMFs near the male speaker's F0. Red: Similarly, for segregation of the female voice, the lowest correlations between (Target + Masker) and Male (Masker alone) occur for IC cells with BMFs in the range of the female F0. B) Solid and dashed: Same as A, but for simulations with SNHL in AN model. Dotted lines: Same as above, except that a simple linear gain of 20 dB was included to increase audibility of speech features in the model population response. (color online)

SUMMARY

This presentation described initial steps towards describing cross-frequency fluctuation contrasts in midbrain responses and their potential for representing sentences in the presence of competing-voice maskers. The fluctuation contrast cues vary across populations of IC neurons with different BMFs, allowing segregation of speech with different F0s. Sensorineural hearing loss reduces the contrasts. Future work will apply these models and correlations to a larger set of sentences for which listener performance data has been collected. The effects of practical amplification schemes on the competing-voice maskers can also be explored for models with different audiometric configurations.

Further work is required to quantitatively evaluate the correspondence between changes in fluctuation contrast cues and changes in listener performance with competing voice maskers. This work could also be extended to studies of competing voices with similar F0s (same gender; e.g., Bramslow *et al.*, 2017).

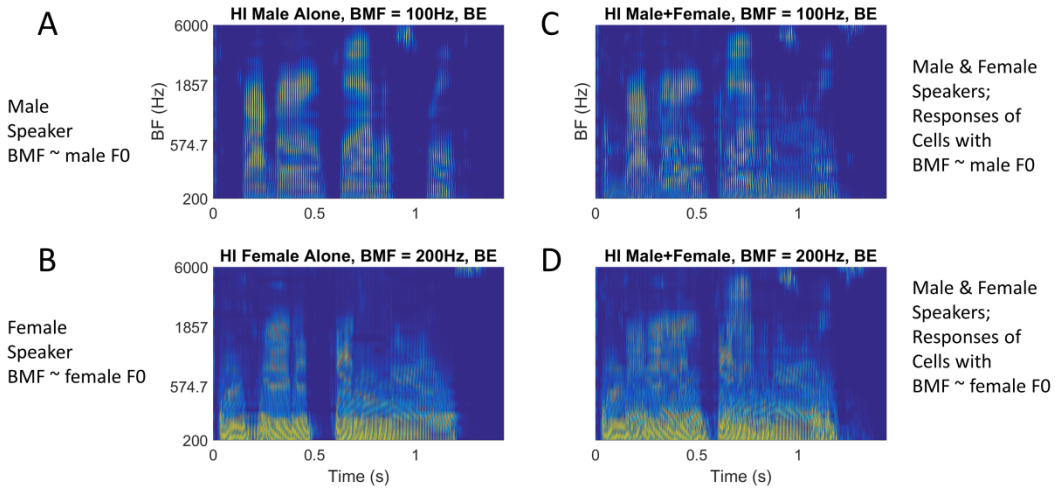

Fig. 3: Examples of population IC model responses to HINT sentences. Model responses here received inputs from a population of HSR AN fibres with sensorineural hearing loss (sloping loss with thresholds elevated by ~10 dB at low CFs to ~40 dB at high CFs). HINT sentences are the same as in Fig. 1. A) Responses of a 100-Hz BMF IC model population to male sentence. B) Responses of a 200-Hz BMF IC population to a female sentence. C) 100-Hz BMF population response to simultaneous sentences. D) 200-Hz BMF population response to the simultaneous sentences. (color online)

ACKNOWLEDGEMENTS

This study was initiated during a sabbatical stay as a Visiting Professor in the Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark. Supported by NIH grants DC001641 and DC010813, and by a fellowship from the Hanse Wissenschaftskolleg, Delmenhorst, Germany.

REFERENCES

Bramsløw, L., Vatti, M., Hietkamp, R.K., and Pontoppidan, N.H. (**2014**). "Design of a competing voices test," Poster presented at International Hearing Aid Conference (IHCON).

Bramsløw, L., Vatti, M., Rossing R., and Pontoppidan, N.H. (**2017**). "An improved competing voices test for test of attention," Proc. ISAAR, **6**, 279-286.

Carney, L.H., Li, T., and McDonough, J.M. (**2015**). "Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations," Eneuro, **2**, ENEURO-0004.

Carney, L.H., Kim, D.O., and Kuwada, S. (**2016**). "Speech coding in the midbrain: Effects of sensorineural hearing loss," in *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing* (Springer), Adv. Exp. Med. Biol., **894**, 427-435. PMID: 27080684

Delgutte, B., and Kiang, N.Y. (**1984**). "Speech coding in the auditory nerve: I. Vowel-like sounds," J. Acoust. Soc. Am., **75**, 866-878.

Joris, P.X., and Yin, T.C. (**1992**). "Responses to amplitude-modulated tones in the auditory nerve of the cat," J. Acoust. Soc. Am., **91**, 215-232.

Joris, P.X. (**2003**). "Interaural time sensitivity dominated by cochlea-induced envelope patterns," J. Neurosci., **23**, 6345-6350.

Joris, P.X., Schreiner, C.E., and Rees, A. (**2004**). "Neural processing of amplitude-modulated sounds," Physiol. Rev., **84**, 541-577.

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

Krishna, B.S., and Semple, M.N. (**2000**). "Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus," J. Neurophysiol., **84**, 255-273.

Langner, G., and Schreiner, C.E. (**1988**). "Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms," J. Neurophysiol., **60**, 1799-1822.

Mao, J., Vosoughi, A., and Carney, L.H. (**2013**). "Predictions of diotic tone-in-noise detection based on a nonlinear optimal combination of energy, envelope, and fine-structure cues," J. Acoust. Soc. Am., **134**, 396-406.

Miller, R.L., Schilling, J.R., Franck, K.R., and Young, E.D. (**1997**). "Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers," J. Acoust. Soc. Am., **101**, 3602-3616.

Nelson, P.C., and Carney, L. H. (**2004**). "A phenomenological model of peripheral and central neural responses to amplitude-modulated tones," J. Acoust. Soc. Am., **116**, 2173-2186. PMCID: PMC1379629

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

Nielsen, J.B., and Dau, T. (**2009**). "Development of a Danish speech intelligibility test," Int. J. Audiol., **48**, 729-741.

Nilsson, M., Soli, S.D., and Sullivan, J.A. (**1994**). "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am., **95**, 1085-1099.

Zilany, M.S.A., Bruce, I.C., and Carney, L.H. (**2014**). "Updated parameters and expanded simulation options for a model of the auditory periphery," J. Acoust. Soc. Am., **135**, 283-286. PMCID: PMC3985897