

# Physiological evidence for auditory modulation filterbanks: Cortical responses to concurrent modulations

**Juanjuan Xiang<sup>a)</sup>**

*Department of Electrical and Computer Engineering, University of Maryland,  
College Park, Maryland 20815  
juanjuan.xiang@gs.com*

**David Poeppel**

*Department of Psychology, New York University, New York, New York 10003  
david.poeppel@nyu.edu*

**Jonathan Z. Simon<sup>b)</sup>**

*Department of Electrical and Computer Engineering,  
University of Maryland, College Park, Maryland 20815  
jzsimon@umd.edu*

**Abstract:** Modern psychophysical models of auditory modulation processing suggest that concurrent auditory features with syllabic ( $\sim 5$  Hz) and phonemic rates ( $\sim 20$  Hz) are processed by different modulation filterbank elements, whereas features at similar modulation rates are processed together by a single element. The neurophysiology of concurrent modulation processing at speech-relevant rates is here investigated using magnetoencephalography. Results demonstrate expected neural responses to stimulus modulation frequencies; nonlinear interaction frequencies are also present, but, critically, only for nearby rates, analogous to “beating” in a cochlear filter. This provides direct physiological evidence for modulation filterbanks, allowing separate processing of concurrent syllabic and phonemic modulations.

© 2013 Acoustical Society of America

**PACS numbers:** 43.64.Qh, 43.64.Ri, 43.66.Ba [BLM]

**Date Received:** September 20, 2012    **Date Accepted:** November 15, 2012

## 1. Introduction

Natural sounds, including animal vocalizations and human speech, are often characterized by the nature of their temporal envelopes. The most critical information for speech intelligibility is preserved in the slowest envelope components, at rates well below 20 Hz (Drullman *et al.*, 1994; Shannon *et al.*, 1995). Phase-locked neural responses to temporally modulated stimuli in human sensory cortex can be noninvasively examined by electroencephalography (EEG) and magnetoencephalography (MEG). Such EEG and MEG signals, when evoked by stationary modulated sounds can be characterized by the auditory steady state response (aSSR), the response component at the same frequency as the stimulus modulation frequency (e.g., Wang *et al.*, 2012). Speech typically contains *multiple* concurrent modulations, but EEG and MEG studies of concurrent modulations have typically focused on rates far above 20 Hz (Lins and Picton, 1995; John *et al.*, 1998; Draganova *et al.*, 2002).

---

<sup>a)</sup>Present address: Goldman Sachs, New York, NY 10282.

<sup>b)</sup>Author to whom correspondence should be addressed. Also at: Dept. of Biology, University of Maryland, College Park, MD 20815.

Two broad categories of theories have been proposed to explain auditory modulation perception. Earlier approaches proposed that the demodulation of input signals is induced by half-wave rectification and compressive processes occurring at the periphery. A low-pass filter in subsequent auditory stages additionally accounts for the observation that a subject's threshold for detecting modulation decreases with increased modulation rates (Viemeister, 1979). A second scheme adds a centrally located bank of bandpass filters that are sensitive to different ranges of modulation frequency (Dau *et al.*, 1997a,b; Jepsen *et al.*, 2008) (see also Chi *et al.*, 1999). This bank of band-limited modulation filters may be thought of as analogous to the cochlear filterbank, but where modulations are segregated by band-limited modulation-filtering, as opposed to the band-limited carrier-filtering of the cochlea.

The present study addresses two questions. First, how are concurrent amplitude modulations physiologically represented in the auditory cortex? Secondly, how do the neural responses to the concurrent modulations fit into theories of modulation filters? We employ sinusoidally amplitude-modulated stimuli containing both single and concurrent modulations (with either a single narrowband or single broadband carrier), at different separations of modulation rate. The concurrent modulations are additive rather than multiplicative (cf. Ewert *et al.*, 2002), so modulation-interaction components are absent at the level of the stimulus. Nonetheless modulation-interaction components may appear in the responses, if the filter outputs only undergo some (unspecified) nonlinearity. This is analogous to the phenomenon of "beating" arising in cochlear filterbank processing of concurrent carriers when nearby enough to be captured by the same cochlear filter. Under this mild assumption, the presence, or absence, of response modulation-interaction components can be used to differentiate between the two types of models: a nonlinear response interaction term (at the frequency given by the difference, or sum, of the frequencies physically present in the stimulus) is evidence that the modulations are processed in the same modulation filter. In contrast, the absence of a nonlinear response interaction term is consistent with the hypothesis that the modulations are processed separately, by distinct modulation filters (Fig. 1).

## 2. Methods

Sixteen subjects (7 males; mean age 24 years) participated in this MEG study. All subjects were right handed (Oldfield, 1971) and had normal hearing and no history of a neurological disorder. The experiments were approved by the University of Maryland Institutional Review Board, and written informed consent was obtained from each participant. Subjects were paid for their participation.

The stimuli, generated using MATLAB (MathWorks Inc., Natick, MA), were 50.25 s in duration with 15 ms onset and offset cosine ramps and were sampled at 44.1 kHz. Three types of conditions were employed: a *single* AM condition (stimulus AM envelope with a single frequency  $f_1$ ), a *nearby* AM-AM condition (stimulus AM envelope with two frequency components  $f_1$  and  $f_2$ , where  $f_2 - f_1 = 3$  Hz), and a *distant* AM-AM condition (stimulus AM envelope with two frequency components  $f_1$  and  $f_2$ , where  $f_2 - f_1 = 17$  Hz). The envelope for the single AM condition is given by  $y_s^{\text{ENV}}(t) = 1 - \cos(2\pi f_1 t)$  and for the concurrent modulation stimuli by  $y_c^{\text{ENV}}(t) = 1 - [\cos(2\pi f_1 t) + \cos(2\pi f_2 t)]/2$ . The six single AM stimulus envelopes were generated with modulation frequencies of 4, 5, 18, 19, 21, and 22 Hz, to verify response measurability of the absence of a concurrent modulation. The two distant AM-AM stimulus envelopes were created by using 4 and 21 Hz and 5 Hz and 22 Hz, respectively. The two nearby AM-AM stimulus envelopes were made with 18 and 21 Hz and 19 and 22 Hz. Finally, these ten envelopes were each applied to two different carriers: a pure tone at 707 Hz, and 5 octave pink noise centered at 707 Hz, giving a total of 20 stimuli.

Subjects were placed horizontally in a dimly lit magnetically shielded room (Yokogawa Electric Corporation, Tokyo, Japan). Stimuli were presented using Presentation software (Neurobehavioral Systems, Albany, CA). The sounds were delivered to

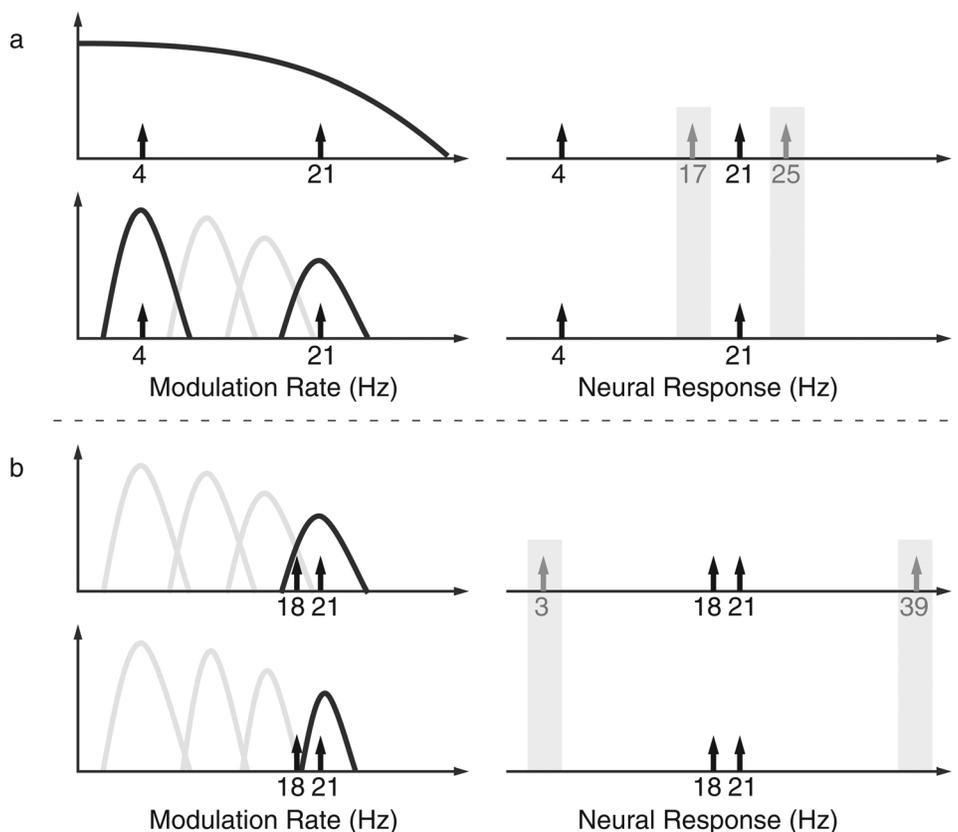


Fig. 1. Cartoon of the effects of auditory modulation filtering on possible resultant neural nonlinearities, as a function of modulation rate and response frequency. (a) Simultaneous presentation of modulations with well-separated rates (e.g., 4 Hz and 21 Hz) should produce responses that include nonlinear interaction terms (e.g., at the difference and sum frequencies of 17 Hz and 25 Hz) if they are processed by the same (broad) neural filter. If they are processed by distinct bandpass filters, however, there would be no within-filter interaction to produce such nonlinear interaction terms. (b) Analogously, simultaneous presentation of modulations of nearby rates (e.g., 18 Hz and 21 Hz) should produce nonlinear interaction terms if they are processed by the same bandpass neural filter (e.g., at the difference and sum frequencies of 3 Hz and 39 Hz), but not if they are processed by still narrower filters.

the subjects' ears with 50  $\Omega$  sound tubing (E-A-RTONE 3A, Etymotic Research, Inc), attached to E-A-RLINK foam plugs inserted into the ear-canal and presented binaurally at a comfortable loudness of approximately 70 dB SPL. Each stimulus was presented once. Interstimulus intervals (ISI) were randomized and ranged uniformly from 1800 to 2200 ms. Subjects listened passively to the acoustic stimuli while MEG recordings were taken.

MEG recordings (157-channel axial gradiometers, KIT, Kanazawa, Japan) were conducted and denoised using the protocols in Xiang *et al.* (2010). For each stimulus, an analysis epoch of duration 50 s (from 0.25 s post-stimulus to the end of the stimulus) was extracted. Each single trial response was transformed using a discrete Fourier Transform (DFT) to a complex frequency response (of 0.02 Hz resolution and 250 Hz extent). The neural responses at 6 modulation frequencies (4, 5, 18, 19, 21, 22 Hz) and 6 potential interaction frequencies (3, 17, 25, 27, 39, 41 Hz) were obtained for each stimulus and channel. The 6 interaction frequencies were further divided into 2 categories, difference rates (obtainable from  $f_2 - f_1$ ) and sum rates (obtainable from  $f_2 + f_1$ ). The remainder of the analysis was based on the normalized neural responses (Xiang *et al.*, 2010), defined as the squared magnitude of the spectral component at

the target frequency divided by the average squared magnitude of the spectral components ranging from 1 Hz below to 1 Hz above the target frequency (excluding the component at the target frequency), averaged over the 20 channels with the strongest individual normalized neural responses.

To assess the potential nonlinearity of the cortical responses to modulations, we used interaction level (IL): the average background-subtracted normalized neural responses at each interaction frequency. The background is estimated to be the average normalized neural response to all stimuli whose envelopes lack a distortion component at this frequency. For example, IL at 3 Hz was calculated by computing the mean normalized neural response at 3 Hz evoked by all the relevant concurrent stimuli (18 and 21 Hz, 19 and 22 Hz), and then subtracting the mean normalized neural response at 3 Hz evoked by all other stimuli. Thus IL is a bias-corrected statistical estimator of the normalized neural response. IL was computed separately for each category: difference rate (3 Hz) vs sum rate (39 Hz, 41 Hz); each modulation condition: nearby vs distant; and each bandwidth: narrowband vs broadband.

### 3. Results

The neural responses to single and concurrent modulated sounds were observed at all direct frequencies (the values of  $f_1$  and  $f_2$  present in the stimulus), with a roughly  $1/f$  power distribution consistent with that seen in Wang *et al.* (2012). The MEG magnetic field distributions of neural responses to single modulations demonstrate the stereotypical patterns of neural activity originating separately from left and right auditory cortex (Elhilali *et al.*, 2009). Similarly, direct neural responses to both of the concurrent modulations emerge as sharp spectral peaks at the individual stimulus component modulation rates  $f_1$  and  $f_2$ , also with stereotypical patterns of neural activity originating separately from left and right hemispheres of auditory cortex.

Neural responses at interaction frequencies ( $f_2 \pm f_1$ ), assessed by IL, were obtained separately for each interaction category (difference frequency vs sum frequencies), each bandwidth (narrowband vs broadband), and each concurrent modulation condition (nearby vs distant). A  $3 \times 2 \times 2$  three-way analysis of variance reveals that carrier bandwidth does not interact with interaction category or modulation condition. Neural responses to stimuli with both narrow and broad bandwidth carriers were therefore pooled together for all further analysis.

We observed that nearby modulation rates produced significant interaction responses, but not distant modulation rates (Fig. 2). The extent of interaction is highly significant for both interaction categories, but, critically, only for the nearby modulation rates and not for distant modulation rates. This is especially striking in the case of the difference frequencies, since the  $\sim 1/f$  spectrum of the background activity (Wang *et al.*, 2012) means the strongest potential to mask detection of the interaction

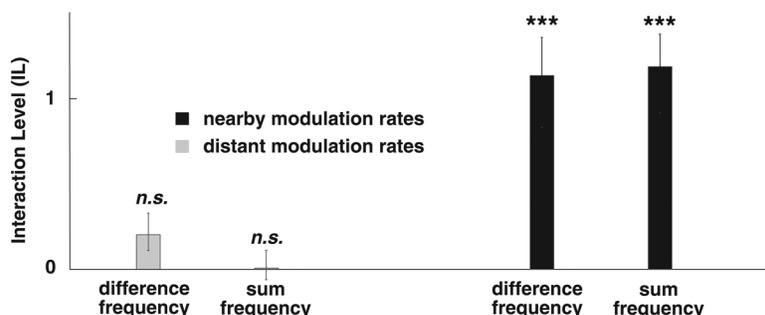


Fig. 2. Interaction Level (IL) by modulation proximity. Distant (well-separated) modulation rates show no evidence of interaction. Nearby modulation rates show highly significant interaction ( $t$  test: \*\*\*  $P < 0.001$ ). Error bars represent one standard error. Responses from both of the two sum-frequencies were pooled together.

frequency occurs for the nearby modulation rates and not for distant modulation rates. This differential activation between the near and distant conditions demonstrates modulation proximity as a critical factor in cortical neural responses to concurrent modulations, suggesting the employment of band-limited modulation filters followed by a nonlinearity.

#### 4. Discussion

The results indicate that the neural response pattern to concurrent modulations depends critically on the rate separation between modulations. The interaction activity indicative of within-channel processing is only evoked from nearby, but not distant, modulation rates, compatible with the physiological employment of central, band-limited modulation filter banks.

Two main categories of modulation filter models have been proposed for the auditory processing of temporal modulation: those containing only peripherally generated lowpass filters (e.g., Viemeister, 1979), and those with additional centrally generated modulation filterbanks (e.g., Dau *et al.*, 1997a,b; Jepsen *et al.*, 2008). Assuming only that that output of the filters is further processed by an (unspecified) nonlinearity, the results here are consistent with filterbank models but not lowpass-only models.

Past studies investigating interaction components of cortical neural responses have not focused on the low modulation rates (near and below 20 Hz) relevant to speech. Lins and Picton (1995) found weak interaction components for concurrent modulations at 81 Hz and 97 Hz. John *et al.* (1998) employed concurrent modulation rates ranging from 70 to 110 Hz with separate carriers and found significant interactions when carrier frequencies were separated by an octave. Draganova *et al.* (2002) investigated neural responses to tones modulated by 38 and 40 Hz concurrently and found a 2 Hz MEG response component. Studies investigating responses to concurrent modulations at the low modulation rates relevant to speech have instead focused on effects of attending to one modulation over the other, rather than on interaction components (Bidet-Caulet *et al.*, 2007; Xiang *et al.*, 2010).

Resolving distant modulation rates in the auditory system is critical for speech perception, since a speech signal can be at least segmented at two time-scales: syllabic rate (near 5 Hz) and phonemic rate (near 20 Hz). The results of this study indicate that the syllabic and phonetic processes are processed independently, but that nearby phonemic rates are processed together.

#### Acknowledgments

Support has been provided by the National Institute for Deafness and Other Communication Disorders (NIDCD) by NIH grants R01 DC 005660 and R01 DC 008342. We thank Mounya Elhilali and Nai Ding for discussions and Jeff Walker for excellent technical support.

#### References and links

- Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P. E., Giard, M. H., and Bertrand, O. (2007). "Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex," *J. Neurosci.* **27**, 9252–9261.
- Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (1999). "Spectro-temporal modulation transfer functions and speech intelligibility," *J. Acoust. Soc. Am.* **106**, 2719–2732.
- Dau, T., Kollmeier, B., and Kohlrausch, A. (1997a). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," *J. Acoust. Soc. Am.* **102**, 2892–2905.
- Dau, T., Kollmeier, B., and Kohlrausch, A. (1997b). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," *J. Acoust. Soc. Am.* **102**, 2906–2919.
- Draganova, R., Ross, B., Borgmann, C., and Pantev, C. (2002). "Auditory cortical response patterns to multiple rhythms of AM sound," *Ear Hear.* **23**, 254–265.

- Drullman, R., Festen, J. M., and Plomp, R. (1994). "Effect of reducing slow temporal modulations on speech reception," *J. Acoust. Soc. Am.* **95**, 2670–2680.
- Elhilali, M., Xiang, J., Shamma, S. A., and Simon, J. Z. (2009). "Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene," *PLoS Biol.* **7**, e1000129.
- Ewert, S. D., Verhey, J. L., and Dau, T. (2002). "Spectro-temporal processing in the envelope-frequency domain," *J. Acoust. Soc. Am.* **112**, 2921–2931.
- Jepsen, M. L., Ewert, S. D., and Dau, T. (2008). "A computational model of human auditory signal processing and perception," *J. Acoust. Soc. Am.* **124**, 422–438.
- John, M. S., Lins, O. G., Boucher, B. L., and Picton, T. W. (1998). "Multiple auditory steady-state responses (MASTER): Stimulus and recording parameters," *Audiology* **37**, 59–82.
- Lins, O. G., and Picton, T. W. (1995). "Auditory steady-state responses to multiple simultaneous stimuli," *Electroencephalogr. Clin. Neurophysiol.* **96**, 420–432.
- Oldfield, R. C. (1971). "The assessment and analysis of handedness: The Edinburgh inventory," *Neuropsychologia* **9**, 97–113.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.* **66**, 1364–1380.
- Wang, Y., Ding, N., Ahmar, N., Xiang, J., Poeppel, D., and Simon, J. Z. (2012). "Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: MEG evidence," *J. Neurophysiol.* **107**, 2033–2041.
- Xiang, J., Simon, J., and Elhilali, M. (2010). "Competing streams at the cocktail party: Exploring the mechanisms of attention and temporal integration," *J. Neurosci.* **30**, 12084–12093.