The Neural Representation of Auditory Modulations Relevant to Speech

Nicholas A. Asendorf and Marisel Villafañe-Delgado
Dr. Jonathan Z. Simon, *Mentor*
Nai Ding and Kai Sum Li, *Graduate Students*

Abstract— **Magnetoencephalography (MEG) is a brain imaging technique that non-invasively measures neurally-generated magnetic fields. Earlier MEG studies have focused on the neural responses to amplitude modulated (AM) auditory signals near 40Hz. Speech signals, however, contain a wide range of modulation rates, most of which are well below 40 Hz. Therefore we seek to characterize the modulation transfer function (MTF) of the human brain at AM frequencies much lower than 40Hz. The present study uses MEG to measure neural responses to pure-tone carrier signals amplitude modulated at frequencies exponentially fluctuating between 3Hz and 60Hz. Analysis of the neural MEG data includes noise reduction, time-frequency analysis to characterize the MTF, and a comparison to the neural response to constant AM stimuli. The maximal neural response was evident at low rate modulations, with the shape of the MTF following that of a shallow low-pass filter. The phase of the neural response was linear, consistent with an 80 ms delay. Neural phase responses to upward and downward sweeps differed by ~π radians for AM frequencies 15-35 Hz. An exponential AM chirp gave a successful estimate of the neural power MTF, closely matching that of the response to constant AM stimuli.**

Index Terms—**auditory, modulation, magnetoencephalography**

I. INTRODUCTION

Speech signals can be described as the mixture of different acoustic components such as the temporal envelope and temporal fine structure [1]. Previous studies show that the temporal envelope is crucial for speech intelligibility, while the temporal fine structure is more important for source localization and pitch perception [2]. Speech envelopes contain a wide range of modulation rates, with most power below 15 Hz. Smith et. al. [3] show that a maximum response of the cortical auditory areas to low frequency amplitude modulated (AM) tones occur between 4 and 16 Hz and these responses match the crucial modulation frequencies for speech intelligibility.

Previous studies [1] suggest that neurons in the auditory cortex are particularly sensitive to the stimulus envelope and can phase-lock to it. They also present a correlation between auditory responses and the temporal envelope of the stimuli. Specifically, they show evidence for phase locked neural activity to AM below 10 Hz.

Magnetoencephalography (MEG) is a non-invasively brain imaging technique that measures neurally generated magnetic fields inside the human brain. Advantages of MEG include high temporal resolution (~1 ms) and moderate spatial resolutions (~1 cm). MEG is advantageous over electro-encephalography (EEG) for studies of auditory neural responses due to the orientation of the auditory cortex, thus the MEG signals are predominantly sensitive to these responses. In addition, magnetic fields have less distortion than electric fields. Previous MEG studies have focused on the neural responses to auditory signals amplitude modulated (AM) near 40Hz [4].

The modulation transfer function (MTF) is commonly used to describe the relation between the neural response and the modulation frequency of the stimulus envelope. This relationship has been observed for some, but not all, modulation frequencies relevant to speech [4]. The main objective of this project is to characterize the power MTF and the phase MTF of the human brain for modulation frequencies lower than 40 Hz, specifically below 15 Hz.

Individual neurons act as frequency-specific filters to certain modulation frequencies, indicated by the MTF of a specific neuron. The MEG signal is a coarse averaging over all auditory areas in the brain. Thus, MEG gives the best non-invasively obtained average MTF of all auditory cortex.

II. METHODS

A. Subjects

Eight volunteers (5 males) participated after providing fully-informed consent. All participants were right handed, had normal hearing, and reported no history of neurological disorders. Among the eight subjects, one subject was excluded from further analysis due to an excess of environmental noise, leaving seven subjects (4 males). The University of Maryland Institutional Review Board approved the experiments.

B. Auditory Stimuli

Nine stimulus conditions were generated using MATLAB (MathWorks Inc, Natick, MA). Each trial was 11 s in duration. Three stimuli had a constant AM rate of 3 Hz, 13 Hz, and 37 Hz with a pure-tone carrier frequency of 707 Hz. The modulation depth was 95%. Three stimuli had an exponentially varying AM rate going from 2.66 Hz to 60.14 Hz with pure-tone carrier frequencies of 250 Hz, 707 Hz, and 2 kHz. Three stimuli had an exponentially varying AM rate

---

This material is based upon work supported by the National Science Foundation under Grant No. 0755224.

Nicholas A. Asendorf is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: asendorf@ umd.edu).

Marisel Villafañe-Delgado is with the Department of Electrical and Computer Engineering, University of Puerto Rico, Mayagüez Campus (e-mail: marisel.villafane@upr.edu).

going from 89.29 Hz to 3 Hz with pure-tone carrier frequencies of 250 Hz, 707 Hz, and 2 kHz. For all stimuli, only the response to the last 10 s of the stimuli was analyzed, as the first second was omitted for an onset response. This gave an AM range of 3 Hz to 60.14 Hz for all exponential sweeps, as the downward sweep was the time reversal of the upward sweep. All stimuli with exponentially changing AM will be further referred to as "chirps". For all six chirp conditions the modulation depth was 95%. The modulation rates are provided below for reference:

$$rate_{upward} = 2 + 3^{0.37t} \qquad \left(-1 \le t \le 10s\right) \tag{1}$$

$$rate_{downward} = 2 + 3^{0.37(10-t)} \qquad \left(-1 \le t \le 10s\right) \tag{2}$$

The stimulus envelope is referred to as $x(t)$ and its Fourier transform is given by $X(\omega)$.

C. *Experimental Procedure*

Each of the exponential chirp stimuli was presented 20 times to the subjects while each of the three constant AM condition types was presented 10 times. The total 180 stimuli were divided into ten blocks of eighteen stimuli. Stimuli were presented to the subject with the ISI (inter stimulus interval) randomly selected from 1.5, 2, and 2.5 s. The subject initiated the progression from one block to the next with a button-press. Subjects were allowed to rest after each block, while required to stay still. The entire experiment took approximately one hour.

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 signals were delivered to the subjects' ears with 50 Ω sound tubing (E-A-RTONE 3A, Etymotic Research, Inc), attached to E-A-RLINK foam plugs inserted into the ear-canal, and presented at a comfortable loudness of approximately 70 dB SPL. The entire acoustic delivery system is equalized to give an approximately flat transfer function from 40-3000 Hz.

Before the main experiment, a pre-experiment was run, where a 707 Hz, 50 ms tone pip was presented 100 times. All pips had a 10 ms onset and offset ramp. The time between pips was randomly selected from 1.5, 2, and 2.5 s and subjects were instructed to count the tone pips. The aim of this task was to record the M100 response (a prominent peak approximately 100 ms after pip onset, also called N1m) to be used for differential source localization.

MEG recordings were conducted using a 160-channel whole-head system (Kanazawa Institute of Technology, Kanazawa, Japan). Its detection coils are arranged in a uniform array on a helmet-shaped surface of the bottom of the dewar, with about 25 mm between the centers of two adjacent 15.5 mm diameter coils. Sensors are configured as first order axial gradiometers with a baseline of 50 mm; their field sensitivities are 5 fT/√Hz or better in the white noise region. Three of the 160 channels are magnetometers separated from the others and used as reference channels in the noise filtering methods. A 200-Hz low-pass filter and a notch filter at 60 Hz were applied to the signal. Two denoising techniques were applied off-line: TS-PCA [5], which removes external noise (filtered versions of the reference channel signals), and SNS [6], which removes noise arising internally from individual gradiometers. TS-PCA was used with a ±100 ms range of time-shifts (filter taps); SNS was used with 10 channel neighbors to exclude sensor noise. Finally, DSS [7], a blind source separation technique designed to preserve phase locked neural activities, was applied. The DSS components are sorted based on how much percent of the response power is phase locked to the stimulus. Only the first DSS component is kept for further analysis in this study.

All data analyses were performed offline in MATLAB after the experimental recordings were completed.

D. *Constant AM Data Processing*

The DFT was computed for the remaining 10 s constant AM observed response (giving a frequency resolution of 0.1 Hz) and averaged over all trials for each subject. The complex magnetic field strength is given by the product of the value of DFT and the sampling interval (1/*fs*), and has units of fT/Hz. [8].

The observed response $y(t)$ is composed of the neural signal, $s(t)$, and background noise, $n(t)$.

$$y(t) = s(t) + n(t) \tag{3}$$

F-tests were implemented for each subject to investigate whether the response at a target frequency was significantly stronger than background noise at that frequency. This test was performed based on the fact that the observed responses $y(t)$ only exhibit neural signal at the stimulus frequency, and only environmental noise at other frequencies. Thus, the response to certain stimulus at the frequency of interest, same as the stimulus modulation rate, was compared to the average of the responses to the other two stimuli at that frequency, $\bar{P}noise$, where only was the noise signal $n(t)$.

If the response at a certain frequency was statistically significant, its power was given by:

$$Power = \left|Y(\omega)\right|^2 - \bar{p}_{noise} \tag{4}$$

where $Y(\omega)$ is the Fourier transform of $y(t)$. In addition to F-tests analysis, ANOVA analyses were done to investigate if the power at each target frequency over subjects has the same distribution.

The statistical calculations performed for the phases were completed following the formulas given by [9] for circular data. The mean phase at each target frequency, $\bar{\theta}$, over trials i, is given by:

$$C = \sum_{i=1}^{n} \cos\theta_i \; S = \sum_{i=1}^{n} \sin\theta_i$$

$$\bar{\theta} = \begin{cases} \tan^{-1}\left(S/C\right) & S > 0,\, C > 0 \\ \tan^{-1}\left(S/C\right) + \pi & C < 0 \\ \tan^{-1}\left(S/C\right) + 2\pi & S < 0,\, C > 0 \end{cases} \tag{5}$$

The mean resultant length $\bar{R}$ for a specific target frequency is:

$$\bar{R} = \sqrt{C^2 + S^2} \tag{6}$$

The sample circular variance *V is*,

$$V = 1 - \bar{R} \tag{7}$$

The sample circular standard deviation *v* is defined by

$$v = \sqrt{-2\log(1 - V)} \tag{8}$$

Significance tests for the phase of the neural response at target frequencies for each subject were completed using permutation methods on the mean resultant length [9]. The first sample, s1, is composed of 9 trials at the target frequency for the response at that frequency, while the second sample, s2, is of size 18, containing the phases of the responses of the other two stimuli at the same target frequency. The observed mean direction resultant length is the difference of the mean resultant length of s1 minus the mean resultant length of s2. The subsequent mean resultant lengths were calculated 10000 times, and then sorted in ascending order. The null hypothesis, that samples came from the same probability distributions, is rejected if the observed mean resultant length is not contained in the lower 95 % of the permutation values. In that case, the hypothesis is rejected at the 5 % significance level.

The phase difference between the response and the stimulus is then is given by:

$$Phase = \bar{\theta}\big(\angle\big(Y(\omega)\big) - \angle\big(X(\omega)\big)\big) \tag{9}$$

The circular standard error of the mean for the phase difference over all subjects is calculated using bootstrap analysis. 1000 different bootstrap re-samplings of the mean phase difference was taken and the circular standard error of the mean was calculated from these 1000 values by:

$$\hat{\sigma}_\omega = v\left(\bar{\theta}_{\mathrm{B}}\right) \tag{10}$$

where $\bar{\theta}_{\mathrm{B}}$ is the phase difference between the response and the stimulus at each re-sample.

E. AM Chirp Data Processing

Data for the twenty trails for each stimulus condition were averaged in the time domain, producing six average responses (one per stimulus condition) for each subject.

A spectrogram was taken for each averaged condition response as well as the envelope of the chirp for that condition. All spectrograms used a 1s Hamming window with 50% overlap and 200 frequency points per DFT (i.e. 1 s). Each spectrogram analyzed 11 s of the neural response starting 0.5 s before the onset and ending 0.5 s after the offset. Because the overlap was 50%, the spectrogram resulted in the 10 s sweep between 3 and 60.14 Hz.

Based on the spectrogram of the stimulus envelope, an extraction window was created. A threshold (0.05 for the discrete power spectral density in each time bin) was set to include areas of the spectrogram corresponding to the stimulus envelope and to exclude all other areas. This window could then be applied to the spectrograms of neural responses to extract the areas that should correspond to the stimulus tract.

The sum was taken over all values at each frequency band (ω_m) in the extracted spectrograms giving an amplitude estimate at each frequency for both the stimulus and neural response. To estimate the power MTF, the window was then applied to both the amplitude spectrum of the neural response, $Y(\omega_m)$, and the stimulus envelope, $X(\omega_m)$, where m corresponds to a discrete frequency.

To estimate the non-neural noise for each condition, time shifted versions of the window were applied to the power spectrum of the neural response, where no neural responses were supposed to occur. For upward sweeps, the window was shifted 4.5 – 7.5 s in 0.5 s steps and for downward sweeps the window was shifted 3.5 – 6.5 s in 0.5 s steps. Shift ranges were selected to best avoid the neural response and any harmonics. The sum was taken over all values at each frequency band in each extracted spectrogram, giving a noise estimate at each frequency. These estimates were then averaged over all seven shifts of the window to give an estimate of the average noise power, $\bar{p}_{noise}$.

The power MTF was then calculated using the following formula:

$$Power_m = \frac{\sum\left|Y(\omega_m)\right|^2 - \bar{p}_{noise}}{\sum\left|X(\omega_m)\right|^2} \tag{11}$$

To estimate the phase MTF, the window was applied to both the complex spectrum of the neural response and the stimulus envelope. The sum was taken over all values at each frequency band in the extracted spectrograms. The phase MTF was then calculated using the following formula:

$$Phase_m = \angle\left(\frac{\sum Y(\omega_m)}{\sum X(\omega_m)}\right) \tag{12}$$

Power and phase values were then averaged (to better increase reliability) over frequency bands of approximately 3 Hz for low AM rates and 5Hz for high AM rates. This resulted in six power and phase MTF plots for each subject, one for each stimulus condition.

III. RESULTS

A. Neural Response

Fig. 1 shows the neural response for one subject by plotting the magnetic fields around the head. The patterns clearly show a magnetic dipole in each hemisphere of the brain, corresponding to the left and right auditory cortexes. Thus, the neural response to the stimuli is an auditory response. Similar figures were generated for the seven subjects who were further analyzed. All subjects showed a normal auditory neural response.

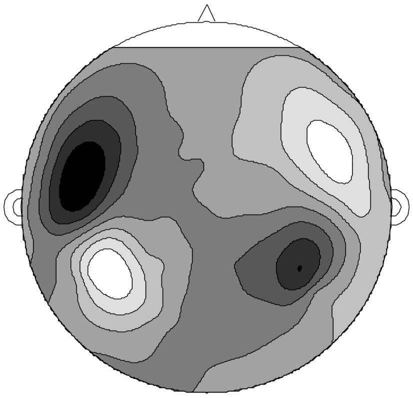

Fig. 1. Distribution of magnetic fields on head. The patterns in each hemisphere represent a magnetic dipole. The sources of the two dipoles are located in the left and right auditory cortexes.

B. Constant AM Response

After F-tests of the response magnitude at each target frequency, it was found that only one subject had no significant response at 13 Hz, $p = 0.44$ ($p < 0.05$ for all other subjects at 13 Hz). All responses were significant at 3Hz ($p < 10^{-3}$) and 37 Hz ($p < 0.02$) for all other subjects.

The average power MTF over subjects (Fig. 2, upper panel) shows a strong response at 37 Hz, but an even higher response is at 3 Hz. The weakest response is at 13 Hz. The variability at 3 Hz is greater and may be because at lower frequencies both the environmental and non-stimulus-driven neural noise is higher.

Tests were performed for the neural power of six subjects at each frequency, after removing the one which response was not significant at 13 Hz. Then, F-tests for power at each frequency over the six subjects show that the neural power came from different distributions for each of the 3 target frequencies. Power at 3 Hz is significantly larger ($p < 4\times10^{-3}$, $F = 14$) than the power at 13 Hz and but not for 37 Hz ($p > 0.1$, $F = 3$). The power at 13 Hz is significantly weaker than the power at 37 Hz ($p < 0.03$, $F = 7$).

Permutation tests show that the phase of the neural responses are significant for all subjects at all tested modulation frequencies, with 95% confidence.

C. AM Chirp Response

Fig. 2 (upper panel) shows the chirp generated power MTF as a dashed line. No differences in power were observed between carrier frequency or direction of the sweep and thus the responses for all six stimulus conditions were averaged for each subject. After averaging over all subjects, the average power MTF shows the strongest neural response at low rate modulations. From modulation rates between 10-40 Hz, the shape of the power MTF is relatively flat. Above 40 Hz, the power of the neural response decreases significantly. For all subjects and conditions, the neural signal was statistically significant compared to background noise.

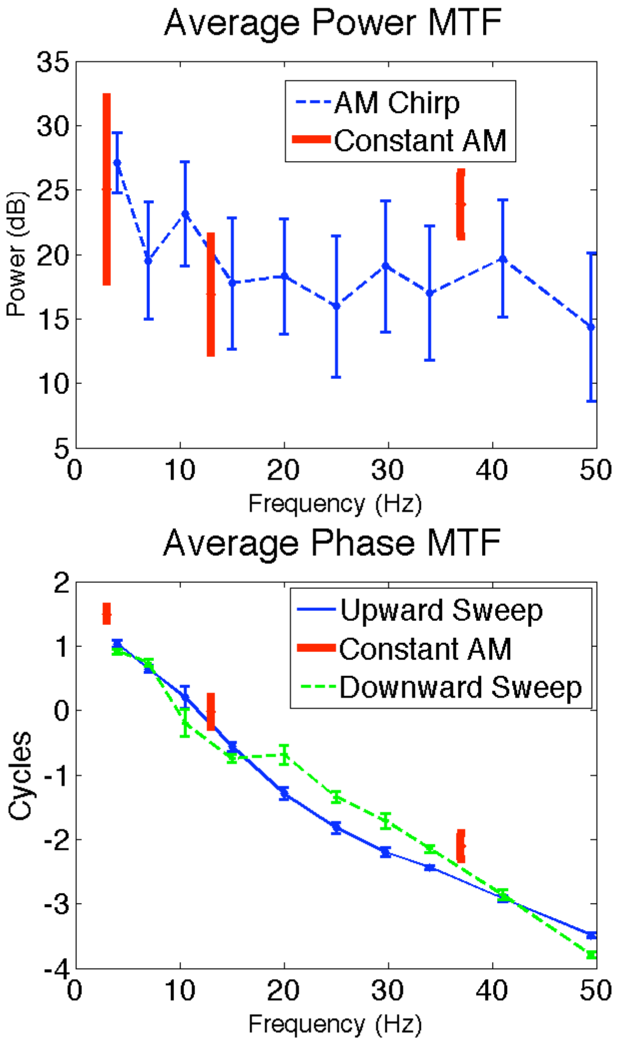

Fig. 2. Power (upper panel) and phase (lower panel) MTF averaged over all subjects. Power error bars are ±2 × standard error and phase error bars are ±2 × standard circular error. For the power MTF, the AM chirp response is averaged across all six stimulus conditions. For the phase MTF, the AM chirp response is averaged over the three upward and downward sweeps respectively. Both the power and phase MTF closely match the response to the constant AM stimuli. The power MTF is strongest at low modulation rates. The phase MTF is linear for both sweep directions. Upward and downward sweeps differ by approximately ½ cycle for modulation rates between 15 and 35 Hz.

Fig. 2 (lower panel) also shows the chirp generated phase MTF. No differences in phase were observed between carrier frequencies and thus the responses were averaged over all upsweeps and downsweeps respectively. Both downward (dashed line) and upward (solid line) sweeps exhibit a linear behavior with a negative slope consistent with an 80 ms delay. From AM rates between 15 and 35 Hz, upward and downward sweeps show a phase difference of ~½ cycle (π radians). However, for low rate modulations and modulations around 40 Hz, upward and downward sweep phase agree.

IV. DISCUSSION

A. Carrier Frequency and Sweep Direction

In both the power and phase MTF, the carrier frequency had no effect on the neural response. All subjects had similar responses to all three carriers of 250 Hz, 707 Hz, and 2 kHz.

This indicates that the brain may respond similarly to a wide range of speech carriers. Intuitively, this makes sense because our ears are trained to listen to many variants of speech on a daily basis and thus a wide range of carrier frequencies would elicit a similar response.

Although sweep direction had no effect on the power MTF, there was a ~½ cycle (π radians) phase difference between upsweep and downsweep neural response for AM rates between 15 and 35 Hz. This may indicate that the pitch inflection may effect how we process speech.

B. *AM Chirp vs. Constant AM*

When the neural response to the exponential AM chirp is compared to the neural response to constant AM sounds, the chirp is a good approximation for both power and phase MTF. For the power MTF, it closely matches for 3 Hz and 13 Hz, where the exponential chirp was slower. For the response around 40 Hz, the difference is ~5 dB (not statistically significant) which may occur because the exponential chirp is quite fast around this frequency range. For the phase MTF, the AM chirp response closely matches at 3 and 13 Hz for both upward and downward sweep conditions. Finally, for the 37 Hz the constant AM response is not consistent with either the upward or downward sweep. This may occur because the chirp does not spend equal time in each frequency range. At lower modulation rates, there are more spectrogram bins that include the neural response than at higher modulation rates. Averaging over more bins produces a more accurate measurement with smaller error. Thus, the chirp response most closely matches the constant AM response at lower frequencies.

C. *Power MTF*

The power MTF shows that the human brain maximally responds to low rate modulations, especially around 3-5 Hz. This behavior is evident in both the AM chirp and constant AM responses. The overall shape of the power MTF is that of a shallow low pass filter. This further indicates that the low rate modulations are the most important in speech, as these would be passed through such a filter. The majority of the power MTF is relatively constant, indicating that no preference is given to frequencies between 10 and 40 Hz.

D. *Phase MTF*

The phase MTF is roughly linear for both sweep directions, with a negative slope consistent with an 80 ms delay. This delay matches previous studies [10] [11], but is longer than delays of studies using separate constant rate modulations [4] [12], especially around 40 Hz. This difference in delay time between AM chirp and constant AM responses may similarly result because the chirp spends a relatively short amount of time at higher frequencies and has less spectrogram bins over which to average. Thus, the phase MTF may actually be different for chirp and constant AM stimuli.

Between AM rates of 15 and 35 Hz, the upward and downward sweep responses differ by approximately ~½ cycle (π radians). However, at low-rate modulations and at modulation rates around 40 Hz, upward and downward sweeps agree in phase. As shown by both [10] [11], the signal-to-noise ratio (SNR) at 40 Hz is the largest. Our results show further evidence of this fact as upward and downward sweeps agree in phase at modulation rates around 40 Hz. Similarly, as the phase for both upward and downward sweeps agree at low modulation rates, it may suggest that the auditory system has a similar strong response at modulations important in speech.

V. CONCLUSION

Compared to constant AM responses, an exponential AM chirp gives a successful estimate of the power MTF and a close estimate of the delay in neural response, especially around low rate modulations. It is also confirmed that the brain responds maximally to low rate modulations, coinciding with frequencies relevant to speech. This is evident as the power MTF behaves like a shallow low-pass filter. The phase MTF of the neural response is linear and is consistent with an 80 ms delay. The phase of neural responses of upward and downward chirp differs by ~½ cycle for modulation rates between 15 Hz and 35 Hz, but agree for low rate modulations and those near 40 Hz.

ACKNOWLEDGMENT

The authors would like to thank Max Ehrmann for his assistance in data collection and the MERIT program and staff for their research support.

REFERENCES

[1] J. Xiang, Y Wang, and J. Z. Simon, "MEG Responses to speech and stimuli with speechlike modulations," *IEEE Proceedings. Neural Engineering*, pp. 33-36, March 16-19, 2005.

[2] C. Liégeois-Chauvel, C.Lorenzi, A. Trébuchon, J. Régis, and P. Chauvel, "Temporal Envelope Processing in the Human Left and Right Auditory Cortices", *Cerebral Cortex,* vol 14, pp 731-740, March 28, 2004.

[3] Z. Smith, B. Delgutte, and A.J. Oxenham, "Chimaeric sounds reveal dichotomies in auditory perception", *Nature*, vol 416, pp. 87-90, Mar. 2002.

[4] B. Roß, C. Borgmann, R. Draganova, L. E. Roberts, and C. Pantev, "A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones", *J. Acoustics. Society of America,* vol. 108, no. 2, pp. 679-691, Aug. 2000.

[5] A. de Cheveigné, and J.Z. Simon, "Denoising based on time-shift PCA" *J. Neuroscience Methods,* vol. 165, no. 2, pp. 297-305, September 30, 2007.

[6] A. de Cheveigné, and J.Z. Simon, "Sensor noise suppression", *J. Neuroscience Methods,* vol. 168, no. 1, pp. 195-202, February 15, 2008.

[7] A. de Cheveigné, and J.Z. Simon, "Denoising based on spatial filtering", *J. Neuroscience Methods,* vol. 171, no. 2, pp. 331-339, June 30, 2008.

[8] M. Elhilali, J. Xiang, S. A. Shamma, and J. Z. Simon, "Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene", *PLos Biology*, vol. 7, no. 6.

[9] N. I. Fisher, *Statistical analysis of circular data* Cambridge University Press, 1993.

[10] J. Artieda et al, "Potentials evoked by chirp-modulated tones: a new technique to evaluate oscillatory activity in the auditory pathway", *J. Clinical Neurophysiology,* vol. 115, no. 3, pp. 699-709, Mar. 2004.

[11] D. W. Purcell, S. M. John, B. A. Schneider, and T. W. Picton, "Human temporal auditory acuity as assessed by envelope following responses", *J. Acoustics. Society of America,* vol. 116, no. 6, pp. 3581–3593, Dec. 2004.

[12] Schoonhoven R, et. al. "A whole head MEG study of the amplitude-modulation-following response: phase coherence, group delay and dipole source analysis." *J. Clinical Neurophysiology,* vol. 114, no. 11, pp. 2096-106, Nov. 2003.