

MEG Responses to Speech and Stimuli With Speechlike Modulations

Juanjuan Xiang¹, Yadong Wang² and Jonathan Z. Simon^{1,3,4}

¹Electrical and Computer Engineering Department, ²Cognitive Neuroscience of Language Laboratory,

³Biology Department, ⁴Bioengineering Program

University of Maryland, College Park

{jxiang, jzsimon}@eng.umd.edu

Abstract - Acoustic signals can be decomposed into the product of an envelope and its fine structure. The envelope features are critical to speech recognition. Results from human behavioral and psychophysical experiments indicate that magnetoencephalography (MEG) responses can significantly phase-lock to the temporal envelope of an auditory stimulus, and speech comprehension is correlated with this phase locking. The goal of this study is to explore the properties of neural response to the speech and stimuli with complex temporal envelopes. Five auditory stimuli with the same envelope, generated from natural speech, and the same stimuli but with the envelope artificially sped up were presented to two subjects. Auditory evoked responses were recorded with a whole-head 160-channel MEG system. Independent Component Analysis (ICA), Principle Component Analysis (PCA) and equivalent current dipole fitting were performed to separate and identify the neural sources (and artifacts). The result shows that the location, orientation, and strength of the neural sources are robust across all stimuli, with correlation between auditory responses and the temporal envelope of the stimuli. These results provide support for the premise that neurons in the auditory cortex are particularly sensitive to the envelope and can phase-lock to stimuli with complex temporal and spectrotemporal envelopes.

I. INTRODUCTION

Speech can be parametrically decomposed into its different acoustic features, such as its spectral envelope, temporal envelope, carrier frequencies and pitch. Among these features, the temporal envelope cues play an important role for speech recognition [1][2][3]. Experiments have shown that the presentation of a dynamic temporal pattern in only a few broad spectral regions is sufficient for the recognition of speech. The frequency of most speech information conveyed by the temporal envelope is below 16 Hz.

Some approaches have been attempted to quantitatively relate this perceptual phenomena with underlying physiological responses of the human brain cortex. Liégeois-Chauvel et al. [4] show that in auditory cortex the predominant response is to the lowest AM frequencies (4–16 Hz), which agrees with multiple animal studies. They also show differences in AM sensitivity

across cortical areas and hemispheres, and provide a physiological foundation for claims of functional specialization of auditory areas. Shannon et al. [3] suggest that only global aspects of the temporal and tonotopic pattern of information available on the auditory nerve are required in speech perception. Ahissar et al. [5] demonstrate that auditory magnetoencephalography (MEG) responses significantly phase-lock to the temporal envelope of the stimulus and that speech comprehension is strongly correlated with this phase locking.

These findings provide a support for the claim of possible correlation between neuronal response and auditory stimulus' envelope. Here, to investigate this correlation, MEG was chosen to record cortical activity in human subjects while passively listening to complex acoustic stimuli. MEG is a leading technology to study the location and dynamics of the human auditory neural activity. MEG has both high temporal resolution (~1ms) and moderate spatial resolution (~1cm) [6]. In addition, due to the anatomical location and configuration of auditory cortex, the MEG signals are particularly sensitive to its activity.

Distinct components of the auditory evoked response can be isolated by Principle Component Analysis (PCA), e.g. in [5], or Independent Component Analysis (ICA), a more sophisticated blind source separation technique. ICA is based on the assumption that since neural sources of activity and any artifacts (e.g. heartbeat or eye movements), are anatomically and physiologically distinct processes, the magnetic signals generated by those processes are statistically independent [7].

II. METHODS

A. Acoustic Stimuli

Ten different stimuli were presented (waveforms are shown in Fig. 1). Most of the temporal envelope energy was limited to below 15 Hz (Fig. 2). The first two were the sentence fragment “she washed her darks” (from the TIMIT speech database [8]) spoken by two different speakers (both have similar temporal envelopes). The temporal envelope from stimulus 2 was extracted by low-pass filtering the Hilbert envelope of the signal and used in the third and fourth stimuli, for which carriers of 5 octave wide pink noise (125 Hz to 4kHz) and of a 400 Hz pure tone were modulated by this temporal envelope. Stimulus 5 also had

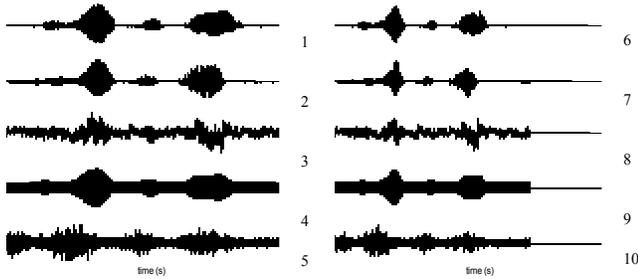


Fig. 1. Waveforms for the 10 stimuli, one second duration shown

same carrier as stimulus 3 but was constructed to have approximately the same spectrotemporal envelope as stimulus 1 & 2. This was done by performing an auditory ripple decomposition [9] of a male speaker whispering the same speech fragment. The ripple decomposition is essentially a two dimensional (spectrotemporal) Fourier decomposition of the spectrogram. Using only the 20 ripples with highest amplitude resulted in a signal which was spectrotemporally low-passed with respect to the original. As a speech fragment it has low (but not zero) intelligibility. All five stimuli were of 1000 ms duration.

To compare the results of this experiment with those of Ahissar et al. [5], the same stimuli were repeated with their envelopes compressed by a factor of 2/3 and their carriers unchanged. A compression algorithm which is based on the phase-vocoder [10] and keeps the spectral and pitch unchanged was applied to stimulus 1 and 2, to produce stimulus 6 and 7 respectively. For stimuli 8, 9 and 10, the envelope was compressed before applying it to the same carriers as stimuli 3, 4 and 5. The last five stimuli were 667 ms in duration.

B. Experimental Conditions

Each stimulus was presented 80 times in pseudorandom order. The interstimulus intervals were randomized and ranged from 800 to 1200 ms. All stimuli were presented binaurally at a comfortable loudness of approximately 70 dB SPL. Two subjects were used (one female). Subjects were native English speakers and right-handed [11], reported normal hearing and no history of neurological disorder. The procedures were approved by the University of Maryland institutional review board and written informed consent was obtained from each participant.

C. Recordings

The magnetic fields were recorded using a 160-channel, whole-head array with superconducting quantum interference device (SQUID)-based first-order gradiometer sensors (KIT, Kanazawa, Japan) in a magnetically shielded room. Three of the 160 sensors were located away from the head and, as reference channels, simply recorded (non-neural) noise. The MEG data were acquired with a sampling rate of 500 Hz, filtered online between 1 Hz and 200 Hz with a notch at 60 Hz. Fast Block LMS adaptive filtering was applied to denoise the 157 neural channels using the three reference channels.

D. Data Analysis

Responses on each channel to each stimulus from 0 ms to 1400 ms post-stimulus-onset were concatenated, resulting in 10 responses for each of the 160 channels.

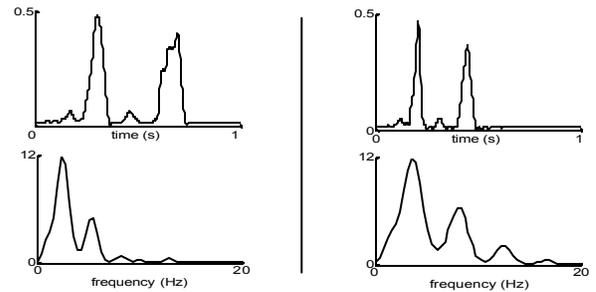


Fig. 2. Envelope of stimuli 2 and PSD of the envelope (left) Envelope of stimuli 7 and PSD of the envelope (right)

ICA (including pre-whitening) was used to decompose the concatenated single-trial MEG data into independent components. In a post-processing stage, these extracted source components were ordered by their spatial contribution on all channels to get the magnetic contour map for the first fifteen powerful source components.

1) Independent Component Analysis

MEG records the superimposed magnetic fields generated by different sources. ICA seeks a linear transformation that maximizes the entropy, since any mixture of independent sources has lower entropy than non-mixed versions, and thus maximizing entropy leads to source separation. Two kinds of sources are usually considered, namely biological sources, such as evoked responses, spontaneous activity (e.g., alpha wave) and artifacts (e.g. eye-blinks and heartbeat), and environmental interference sources, such as 60 Hz electrical power interference [12]. Sources of different natures usually have different high-order moments, and a signal separation algorithm must enable the separation of mixtures of sub-Gaussian and super-Gaussian signals. This is done by an extended infomax ICA algorithm [7]. In implementation, the algorithm starts from an initial (identity) demixing matrix, updates itself according to the maximum entropy principle, and terminates when the directional change is less than 10^{-7} or when change in the directional change less than 10^{-7} (300 iterations maximum). The mixing matrix is the inverse of this demixing matrix.

2) Equivalent-Current Dipole Localization

The (possibly) extended neuronal sources can be represented by single equivalent-current dipoles. The source localization is sought which best minimizes the difference between the solution of the forward model and the measured signals [13].

3) Principle Component Analysis

To compare the results of this experiment with those of Ahissar et al. [5], PCA was performed on the single-trial MEG data as well. The magnetic field distribution of the first principle component (PC1) was plotted, the power spectral density (PSD) of PC1 was compared with the PSD of stimulus' envelope, and the correlation coefficient between PC1 and stimulus envelope was calculated.

III. RESULTS

A. Magnetic Field Distribution of Response Components

The independent components were extracted by applying ICA

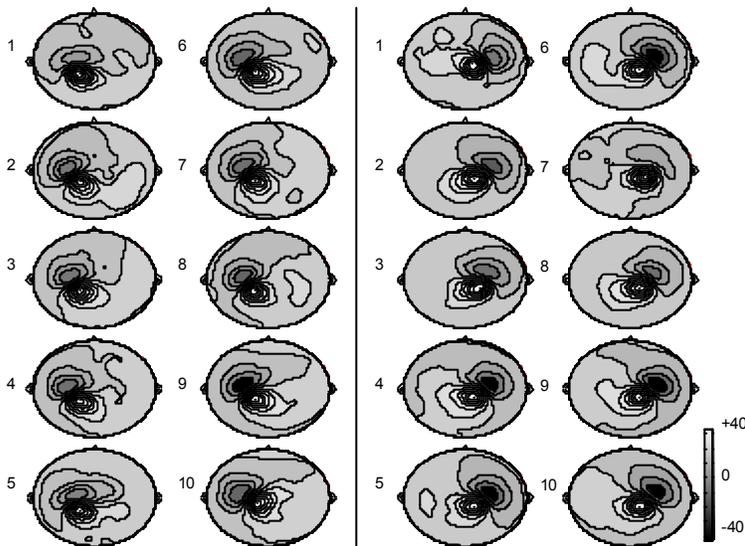


Fig. 3. Smoothed spatial maps for the auditory independent components of the response to all stimuli, for subject 1

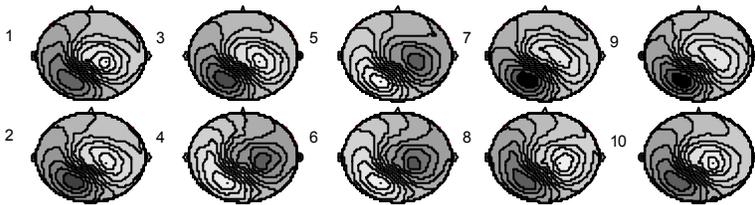


Fig. 4. Smoothed spatial maps for the first principle components of the response to all stimuli, for subject 1

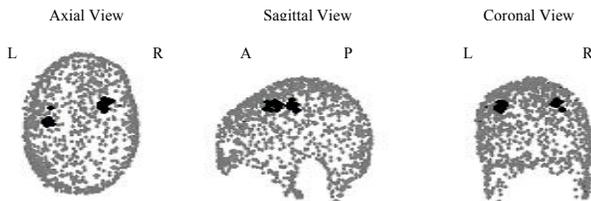


Fig. 5. Equivalent-current dipoles for every left and right auditory IC, for subject 1

to the MEG responses. Each independent component has a time waveform and a spatial map representing its contribution at each sensor on the head. Fig. 3 shows maps of those independent components with support over the left auditory cortical (left aIC) areas in the left column and the corresponding right auditory independent component (right aIC) in the right column. In contrast, the head maps generated by PC1 are shown in Fig. 4.

It was observed, for both subjects, that for every stimulus there is always one distinct independent response component localized over the left auditory cortex and another localized over the right cortex. This pattern is consistent across stimuli. In contrast, the PC1 was always located medially.

The left and right aICs were distinguished from other independent components produced by ICA according to two separate characteristics. The one discussed already is that the spatial map is highly stereotypical, and gives high spatial

correlations with each other (across stimuli) but not with other independent components. The other is a stereotypical PSD. The integrated difference between one auditory IC and other auditory independent components (across stimuli) was smaller for the auditory IC than for any other independent component. These correlations were so strong that there was never *any* ambiguity as to which independent components were the aICs.

A single equivalent-current dipole procedure was performed to estimate the location, orientation and intensity of the dipole which produces the corresponding magnetic field. The results are shown in Fig. 4. The distribution of goodness of fit for the two subjects were $90\pm 8\%$ and $87\pm 7\%$. Despite the low number of trials, the dipoles localize tightly. The standard deviations of the right dipole sources were 6.6 mm and 10.3 mm and 5.2 mm and 9.2 mm for the left dipole (Fig. 5).

For PC1, because the field concentration is so medial, the maps are impervious to biologically plausible dipole fits. Despite the lack of a spatial/anatomical relationship to auditory cortex, however, the time course of the auditory PC has strong physiological relevance, as discussed next.

B. Spectral Overlap and Cross-Correlation

Ahissar et al. [5] demonstrated that both the PSD and the time waveform of the PC1 have strong overlap with the stimulus envelope, for moderately compressed speech. This experiment confirms and expands these results.

First we address the overlap in PSD between the response PC1 and the stimulus envelope, for compressed stimuli. This is demonstrated in the rightmost panel of Fig. 6. There the PSD of the PC1 and the stimulus 7 envelope is shown by solid line and dashed line respectively. This overlap should not be surprising, since the PSD of the envelope is quite typical for speech (expanded moderately to higher frequencies by the compression), and the PSD of the PC1 response is quite typical for an evoked MEG response. For the PC1 response to uncompressed stimuli, the overlap in PSD between the PC1 and the stimulus envelope is also present, though to a lesser extent.

In addition to the spectral power, the spectral phase was also partially preserved. This is most easily seen by the correlation coefficients between the time waveforms of the PC1 and the stimulus envelope, which is shown in the rightmost panel of Fig. 7. Non-trivial values of the correlation coefficient show evidence of phase-locking. For both subject, the peak correlation coefficient for the five compressed stimuli range from 25% to 50% - evidence for phase-locking. The peak correlation coefficient for the five uncompressed stimuli range from 25% to

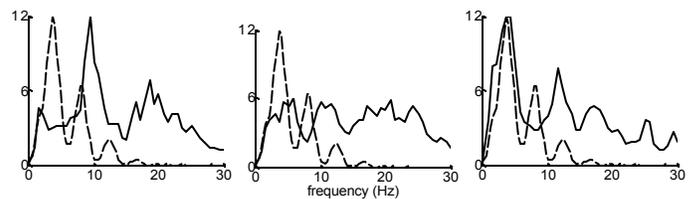


Fig. 6 PSD of responses (left aIC, right aIC, PC1) and stimulus envelope (dashed) for stimulus 7 and subject 1

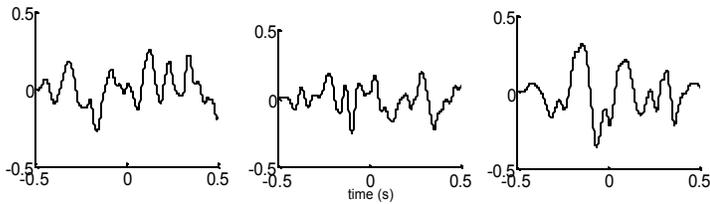


Fig.7 Correlation between responses (left aIC, right aIC, PC1) and stimulus envelope, for stimulus 7 and subject 1

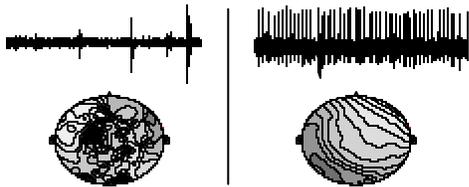


Fig. 8 Time waveform and smoothed spatial maps of IC1 (left), Time waveform and smoothed spatial maps of IC2 (right), for stimulus 3

39% - also evidence for phase-locking, though weaker.

In contrast to the PC1s, the aICs do not show any particular spectral overlap. In Fig. 6, the PSD of the left aIC and the stimulus 7 envelope is showed in left by solid line and dashed line respectively. The PSD of the right aIC and the stimulus 7 envelope is showed in middle by solid line and dashed line. The PSD is typically flat almost all the way up to 30 Hz, occasionally dominated by an additional peak in the alpha band (~10 Hz). This is not at all similar to the PSD of the stimulus envelope, which is strongly peaked well below 10 Hz and falls off quickly by 10 Hz.

Despite this, the peak correlation coefficient for the five uncompressed stimuli still range from 20% to 44%, evidence for some phase-locking for frequencies below 10 Hz. The peak correlation coefficients for the five compressed stimuli similarly range from 17% to 46%. These figures are comparable to the peak correlation coefficients between the left aIC and the right aIC, for each stimulus, which range from 25% to 31% for the five uncompressed stimuli and from 22% to 28% for the five compressed stimuli. The left panel of Fig. 7 shows the correlation coefficient between the left aIC and the stimulus envelope and the right panel shows correlation coefficient between the right aIC and the stimulus envelope.

C. Artifact Identification.

During MEG measurement some undesirable artifacts, e.g. eye activity, heart beat and environmental magnetic disturbances, were recorded simultaneously. The artifacts may have large amplitude, and some artifacts resemble the neural activity, thus leading to a misinterpretation of the data.

ICA is a useful technique to identify and remove these artifacts in MEG. Fig. 7 presents the first independent components (IC1) and the second independent component (IC2), in the response to stimulus 3 for subject 1. Based on the time waveform, IC1 is generated from eye blinks. IC2 shows a cardiac rhythm. For each component, the head maps are also shown.

IV. CONCLUSION

In this paper, we have shown that for a variety of complex auditory stimuli sharing the same temporal envelope, ICA identifies auditory related responses, localized over left and right auditory cortex independently, in addition to extracting artifacts. Furthermore the location and the orientation of auditory dipoles are robust across stimuli.

We also have confirmed that for speech and speech derived signals, the response has a strong spectral overlap with the acoustic temporal envelope. There is significant correlation between not only PCA responses and stimuli but also ICA responses and stimuli. Taken together, our results suggest that neurons in the auditory cortex can phase-lock to stimuli with complex temporal and spectrotemporal envelopes.

REFERENCES

- [1] R. Drullman, J. M. Festen, and R. Plomp, "Effect of Reducing Slow Temporal Modulations on Speech Reception", *J. Acoustic. Soc. Amer.*, 95, 2670-2680, 1994.
- [2] R. V. Shannon, F. Zeng, J. Wygonski, V. Kamath, and M. Ekelid, "Speech Recognition with Primarily temporal Cues", *Science* 270, 303-304, 1995.
- [3] S. Greenberg, T. Arai, "The Relation between Speech Intelligibility and the Complex Modulation Spectrum", *Proceedings of the 7th Eurospeech Conference on Speech Communication and Technology*, 473-476, 2001
- [4] C. Liegeois-Chauvel, C. Lorenzi, A. Trebuchon1, J. Regis and P. Chauvel, "Temporal Envelope Processing in the Human Left and Right Auditory Cortices, *Cerebral Cortex*", Vol: 14, 731-740, 2004.
- [5] E. Ahissar, et al. "Speech Comprehension is Correlated with Temporal Response Patterns Recorded from Auditory Cortex", *Proceeding of the National Academy of Science*. 98(23),13367-13372.
- [6] M. Hamalainen et al., "Magnetoencephalography-Theory, Instrumentation, and Applications to Noninvasive Studies of the Working Humans Brain", *Reviews of Modern Physics*, 65, 413-197, 1993.
- [7] T. Lee, M. Girolami, and T. J. Sejnowski, "Independent Component Analysis using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources", *Neural Computation*, Vol. 11, No. 2, 417- 441, 1999.
- [8] L. F. Lamel, J. Garafolo, W. Fiscus, and D. S. Pallett "TIMIT: The DARPA Acoustic-Phonetic Speech Corpus", NTIS Publication, 1990.
- [9] S. Shamma, "Auditory Cortical Representation of Complex Acoustic Spectra as Inferred from the Ripple Analysis Method", *Network: Computation in Neural Systems*, Vol. 7, 439-476, 1997.
- [10] M. R. Portnoff, "Time Scale Modification of Speech based on Short Time Fourier Analysis", *IEEE Trans. Acoust.* 29, 1374-1390, 1981.
- [11] R.C. Oldfield, The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 1971; 9:97-113.
- [12] J. Cao, N. Murata, S. Amari, A. Cichocki and T. Takeda, "Independent Component Analysis for Unaveraged Single-trial MEG Data Decomposition and Single-dipole Source Localization", 255-277, *Neurocomputing*, Volume 49, 2002.
- [13] Mosher, J. C., Leahy, R. M., Lewis, P. S., "EEG and MEG: Forward Solutions for Inverse Methods", *IEEE Trans Biomed Eng*, 46, 245-259, 1999.