



# Time-domain analysis of neural tracking of hierarchical linguistic structures



Wen Zhang<sup>a,1</sup>, Nai Ding<sup>a,b,c,\*</sup>

<sup>a</sup> College of Biomedical Engineering and Instrument Sciences, Zhejiang University, Hangzhou, China

<sup>b</sup> Interdisciplinary Center for Social Sciences, Zhejiang University, Hangzhou, China

<sup>c</sup> Neuro and Behavior EconLab, Zhejiang University of Finance and Economics, Hangzhou, China

## ARTICLE INFO

### Keywords:

MEG  
Canonical component analysis  
Inter-subject correlation  
Neural entrainment  
Speech

## ABSTRACT

When listening to continuous speech, cortical activity measured by MEG concurrently follows the rhythms of multiple linguistic structures, e.g., syllables, phrases, and sentences. This phenomenon was previously characterized in the frequency domain. Here, we investigate the waveform of neural activity tracking linguistic structures in the time domain and quantify the coherence of neural response phases over subjects listening to the same stimulus. These analyses are achieved by decomposing the multi-channel MEG recordings into components that maximize the correlation between neural response waveforms across listeners. Each MEG component can be viewed as the recording from a virtual sensor that is spatially tuned to a cortical network showing coherent neural activity over subjects. This analysis reveals information not available from previous frequency-domain analysis of MEG global field power: First, concurrent neural tracking of hierarchical linguistic structures emerges at the beginning of the stimulus, rather than slowly building up after repetitions of the same sentential structure. Second, neural tracking of the sentential structure is reflected by slow neural fluctuations, rather than, e.g., a series of short-lasting transient responses at sentential boundaries. Lastly and most importantly, it shows that the MEG responses tracking the syllabic rhythm are spatially separable from the MEG responses tracking the sentential and phrasal rhythms.

## Introduction

In the human language, smaller linguistic units such as syllables and words can be recursively combined into larger linguistic structures such as phrases and sentences. How linguistic units of different sizes are represented in the brain is a fundamental question in cognitive neuroscience (Buiatti et al., 2009; Everaert et al., 2015; Garrett et al., 1966; Pallier et al., 2011; Peña and Melloni, 2012; Townsend and Bever, 2001). It is shown that when listening to continuous speech, cortical activity recorded by magnetoencephalography (MEG) and electroencephalography (EEG) follows the rhythms of acoustic features of speech that are related to syllabic and phonemic level processing (Di Liberto et al., 2015; Ding and Simon, 2012a, b; Gross et al., 2013; Kayser et al., 2015; Kerlin et al., 2010; Luo and Poeppel, 2007). Recently, it is further shown that cortical activity can concurrently follow higher level linguistic structures such as phrases and sentences using speech materials illustrated in Fig. 1a (Ding et al., 2016).

Neural tracking of phrasal and sentential structures, however, was only characterized in the frequency domain by analyzing the global field power, leaving several questions unanswered. First, when cortical

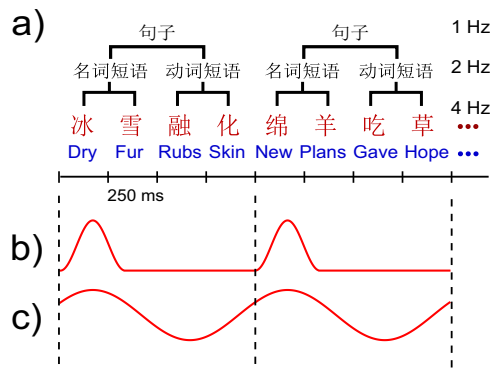
activity become entrained/synchronized to the phrasal and sentential rhythms, how long does it take for entrained activity to reach a steady state? The buildup timecourse of entrained activity depends on at least 2 factors. First, it depends on the dynamic properties of the underlying neural sources. For example, the auditory Steady State Response (aSSR) to a sound repeating at 40 Hz builds up in ~300 ms, after ~12 cycles of the stimulus (Ross et al., 2002). Second, it depends on how much time the brain needs to parse the temporal structure of the input. For example, the aSSR may take up to 4 s to build up when the periodic stimulus is interfered by competing sounds (Elhilali et al., 2009). Here, we employ language materials that are easy to parse to focus on the buildup process constrained by the dynamical properties of the underlying neural network.

Second, frequency-domain analysis does not directly illustrate the response waveform. Therefore, it is not intuitive whether the sentential-rate response continuously changes over the timecourse of a sentence (Fig. 1b) or whether it only shows an abrupt change at sentential boundaries (Fig. 1c). If the neural response is a continuously changing slow oscillation, it could be interpreted as an integrator that accumulate information over the timecourse of a sentence (Pallier

\* Corresponding author at: College of Biomedical Engineering and Instrument Sciences, Zhejiang University, Zhejiang 310027, China.

E-mail address: [ding\\_nai@zju.edu.cn](mailto:ding_nai@zju.edu.cn) (N. Ding).

<sup>1</sup> These authors contributed equally to this work.



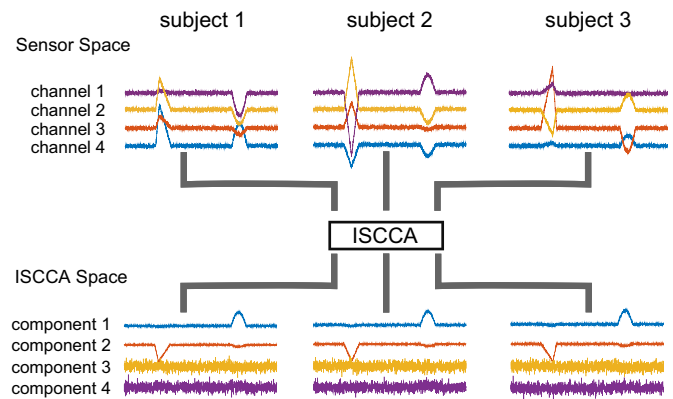
**Fig. 1.** Linguistic structure of the stimulus (a) and possible neural responses (b-c). (a) A sequence of Chinese syllables are presented isochronously at 4 Hz. Examples in English are also shown for illustrative purposes. All syllables are independently synthesized by a computer. Two syllables construct a phrase and two phrases construct a sentence. Therefore, the syllables, phrases, and sentences are presented at 4 Hz, 2 Hz, and 1 Hz respectively. This figure is adapted from Ding et al. (2016). (b-c) Two hypotheses about how cortical activity follows the sentential rhythm, whether it continuously changes over time (b) or occurs only briefly at sentential boundaries (c).

et al., 2011). In contrast, if the neural response is a short-lasting transient response at structural boundaries, it is more appropriately interpreted as a change/boundary detector.

Third, previous frequency-domain analysis was based on the global field power of MEG, leaving it unclear whether the neural tracking of different linguistic levels can be spatially separated using MEG? To answer the above mentioned 3 questions, we apply a time-domain analysis of MEG responses. Furthermore, although previous studies assume that the sentential- and phrasal-rate neural responses are entrained, i.e., phase locked, to the stimulus but the degree of phase locking is not quantified. Here, we quantify the consistency of the neural response phases over subjects by calculating the inter-subject phase coherence (Fisher, 1993).

A time-domain analysis of MEG responses, however, is difficult for several reasons. First, each neural generator, i.e. a current source, produces a source/sink pattern in the MEG signal. The MEG signals from the source and the sink have opposite polarities and the spatial locations of the source and sink patterns are not aligned across subjects due to the anatomical differences and the subjects' head position inside the MEG machine. Second, there are usually multiple neural generators contributing to the neural tracking of a continuous stimulus and these neural generators could have different response phases due to their positions in the neural processing hierarchy or their neurodynamical properties. As a result, if the macroscopic MEG responses are dominated by different neural generators in different subjects due to anatomical differences, they will show phase differences across subjects. Lastly, the MEG signal is a mixture of the responses from multiple neural generators and component analysis methods, such as the principal component analysis (PCA), independent component analysis (ICA), and denoising source separation (DSS), are often employed to separate different neural sources. The polarity of the response waveforms extracted by the component analysis, however, is usually arbitrary, which further increases the difficulty for grand averaging the response waveforms across subjects.

Although it is difficult to align the response phase across subjects in MEG, recent studies have shown that the neural response phase is relevant to perception (Henry and Obleser, 2012; Lakatos et al., 2008; Schroeder and Lakatos, 2009) and shows consistency across subjects during the processing of continuous natural stimuli (Dmochowski et al., 2014; Hasson et al., 2012; Hasson et al., 2004; Honey et al., 2012; Lankinen et al., 2014). For normal listeners, during speech comprehension, it is reasonable to assume and empirical studies support that common neurophysiological processes underlie the processing of the same unambiguous sentence.



**Fig. 2.** Illustration of the basic function of the Inter-Subject Coherent Component Analysis (ISCCA). The ISCCA decomposes the multi-channel recordings from individual subjects into components, by maximizing the correlation between response waveforms across subjects. In this illustration, a 4-channel recording was simulated for 3 subjects. Each recording is a mixture of an early response, a late response, and white noise. The late response has the same waveform across subjects and is captured by the first ISCCA component. The waveform of the early response slightly varies across subjects and is captured by the second ISCCA component. The noise signal is captured by the 3rd and the 4th ISCCA components. The early response is simulated by a sawtooth signal. Its phase is identical within each subject across channels but varies across subjects. Both the early and the late responses have random polarity and amplitude in each channel. In this illustration, since the data have only 4 channels, the DSS dimension reduction step is omitted and the mCCA is applied to the 4-channel data directly.

To optimally extract neurophysiological processes that are common across subjects, we propose an analysis method called the Inter-Subject Coherent Component Analysis (ISCCA). The ISCCA decomposes the multi-channel MEG recordings of each subject into components and maximizes the inter-subject correlation of each component. Each ISCCA component is extracted by a spatial filter and can be viewed as the recording from a virtual sensor spatially tuned to a cortical network that shows coherent neural activity over subjects. The ISCCA spatial filters are subject-specific and normalize individual differences in response topography. Since the ISCCA components are maximally correlated over subjects, they can be directly averaged for group level analysis. As an illustration, Fig. 2 shows that responses may have very different amplitude and polarity in different channels in the sensor space. However, when the responses are projected to the ISCCA space, responses that show coherence over subjects are attributed to the same ISCCA component with the same polarity, which facilitates group-level analysis of the response waveform.

In the following, we apply the ISCCA to extract MEG response components that have a coherent response waveform over subjects and analyze the time course of neural tracking of linguistic structures based on the grand averaged response waveform.

## Materials and methods

### Experimental procedures

Sixteen healthy young adults participated in the experiments and the data analyzed here were previously reported by Ding et al. (2016). In the experiment, the subjects listened to an isochronous sequence of syllables. These syllables were ordered so that neighboring 4 syllables constructed a sentence (Fig. 1a). Each sentence was composed of a noun phrase (2 syllables) followed by a verb phrase (2 syllables). The syllables were presented at a constant rate of 4 Hz and no pause was inserted between phrases or sentences. Therefore, the sentences were presented at 1 Hz and the phrases were presented at 2 Hz.

In each trial, 40 syllables were played and 28 trials were collected. To ensure attention, the subjects were instructed to detect semantically abnormal sentences such as “green frogs drove cars” by a button press at the end of the trial. Eight trials contained abnormal sentences and

Download English Version:

<https://daneshyari.com/en/article/5631345>

Download Persian Version:

<https://daneshyari.com/article/5631345>

[Daneshyari.com](https://daneshyari.com)