NeuroImage 172 (2018) 162–174

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension

Christian Brodbeck [a,*], Alessandro Presacco [b], Jonathan Z. Simon [a,c,d]

[a] Institute for Systems Research, University of Maryland, College Park, MD, USA
[b] Department of Otolaryngology, University of California, Irvine, CA, USA
[c] Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA
[d] Department of Biology, University of Maryland, College Park, MD, USA

ARTICLE INFO

Keywords:
Magnetoencephalography
Minimum norm estimate
Speech representation
Impulse response
Temporal response function
Reverse correlation

ABSTRACT

Human experience often involves continuous sensory information that unfolds over time. This is true in particular for speech comprehension, where continuous acoustic signals are processed over seconds or even minutes. We show that brain responses to such continuous stimuli can be investigated in detail, for magnetoencephalography (MEG) data, by combining linear kernel estimation with minimum norm source localization. Previous research has shown that the requirement to average data over many trials can be overcome by modeling the brain response as a linear convolution of the stimulus and a kernel, or response function, and estimating a kernel that predicts the response from the stimulus. However, such analysis has been typically restricted to sensor space. Here we demonstrate that this analysis can also be performed in neural source space. We first computed distributed minimum norm current source estimates for continuous MEG recordings, and then computed response functions for the current estimate at each source element, using the boosting algorithm with cross-validation. Permutation tests can then assess the significance of individual predictor variables, as well as features of the corresponding spatio-temporal response functions. We demonstrate the viability of this technique by computing spatio-temporal response functions for speech stimuli, using predictor variables reflecting acoustic, lexical and semantic processing. Results indicate that processes related to comprehension of continuous speech can be differentiated anatomically as well as temporally: acoustic information engaged auditory cortex at short latencies, followed by responses over the central sulcus and inferior frontal gyrus, possibly related to somatosensory/motor cortex involvement in speech perception; lexical frequency was associated with a left-lateralized response in auditory cortex and subsequent bilateral frontal activity; and semantic composition was associated with bilateral temporal and frontal brain activity. We conclude that this technique can be used to study the neural processing of continuous stimuli in time and anatomical space with the millisecond temporal resolution of MEG. This suggests new avenues for analyzing neural processing of naturalistic stimuli, without the necessity of averaging over artificially short or truncated stimuli.

Introduction

In a natural environment, the brain frequently processes information in a continuous fashion. For example, when listening to continuous speech, information is extracted incrementally from an uninterrupted acoustic signal at multiple levels: phonetically relevant sound patterns are recognized and grouped into words, which in turn are integrated into phrases which are meaningful in the context of a larger discourse (e.g. Gaskell and Mirkovic, 2016). Contrary to this continuous mode of functioning, neuroimaging experiments typically isolate phenomena of interest with short, repetitive trials (for many examples, see e.g. Gazzaniga et al., 2009). While such research unquestionably leads to valuable results, the lack of naturalness of the stimuli is associated with uncertainty of how generalizable such results are to real world settings (see e.g. Brennan, 2016). Consequently, there is a need for complementary research with more naturalistic stimuli.

Brain responses to continuous speech have been studied with functional magnetic resonance imaging (fMRI) (Brennan et al., 2012; Brennan et al., 2016; Chow et al., 2014; Willems et al., 2016). Hemodynamic changes have been shown to track inherent properties of words, such as

* Corresponding author.
E-mail address: brodbeck@umd.edu (C. Brodbeck).

https://doi.org/10.1016/j.neuroimage.2018.01.042
Received 9 September 2017; Received in revised form 12 December 2017; Accepted 17 January 2018

word frequency, as well as properties of words in context, such as the contextual probability of encountering a given word. However, the low temporal resolution of fMRI, typically sampled at or below 1 Hz, imposes several limitations on the phenomena that can be modeled. While the studies cited above suggest that the resolution is adequate to model responses with a timescale of individual words, this is not the case for processes at faster timescales such as phonetic perception, where relevant events last only tens of milliseconds. In addition, fMRI responses can be modeled in terms of brain regions which are or are not sensitive to a given variable, but the relative and absolute timing of different components of the response remain obscure. Thus, even when word-based variables are analyzed, hemodynamic responses are modeled as instantaneous effects of the relevant variable, convolved with the hemodynamic response function, but without taking into account the temporal relationship between the stimulus and different components of the brain response (e.g. Brennan et al., 2016; Willems et al., 2016).

In contrast to fMRI, electroencephalography (EEG) and magnetoencephalography (MEG) have the temporal resolution to track continuous processing with millisecond accuracy. Previous research has established that the dependency of the MEG or EEG response on a continuous stimulus variable can be modeled as a linear time-invariant system (Lalor et al., 2006). This technique has been originally developed for relating neurons' spiking behavior to continuous sensory stimuli (see Ringach and Shapley, 2004), but can be extended to MEG/EEG signals by modeling the response as a linear convolution of a stimulus variable with an impulse response function (see Fig. 1). Given a known stimulus and a measured response, one can then estimate the optimal response function to predict the measured response from the stimulus. This technique has been used to model EEG responses to continuously changing visual stimuli, by modeling continuous EEG signals as the convolution of moment-by-moment stimulus luminance with an appropriate response function (Lalor et al., 2006). An analogous procedure has been used to estimate responses to amplitude modulated tones and noise (Lalor et al., 2009). As an extension of this procedure, the response to continuous speech has been modeled as a response to the level of momentary acoustic power, the acoustic envelope (Lalor and Foxe, 2010).

While the original formulation focused on purely sensory neurons, i.e. neurons whose response is a linear function of sensory input (Ringach and Shapley, 2004), the same method has also been applied successfully to determine cognitive influences on sensory processing. This can be achieved by modeling the signal as a response to a continuous predictor variable that represents a specific property of interest of the input stimulus. Thus, besides the acoustic envelope, the EEG response to continuous speech has been shown to reflect categorical representations of phonemes (Di Liberto et al., 2015). Furthermore, using stimuli in which speech from multiple talkers is mixed, it has been shown that the response function to the acoustic envelope can be divided into an earlier component around 50 ms that responds to the acoustic power in the overall stimulus, and a later component around 100 ms that responds to the acoustic envelope of the attended speech stream but not the unattended one (Ding and Simon, 2012a,b).

While this research shows that response functions for continuous stimuli can be estimated, and that they can track not just sensory but also cognitive processes, all the above studies estimated response functions using only sensor space data. Topographic distributions of response functions have been assessed using equivalent current dipole localization (Lalor et al., 2009; Ding and Simon, 2012a) but this does not use the full localizing power of MEG. For investigating cognitive processing of sensory signals in particular, better source localization has the potential to separate response functions to different stimulus properties through anatomical separation of the brain response. In this paper, we propose to use distributed minimum norm source estimates to localize MEG data before estimating response functions. We developed a procedure in which source estimates are computed for continuous raw data, response functions are estimated independently at each virtual current dipole of the source model, and these individual response functions are then recombined to create a picture of the brain's responses to different functional aspects of the continuous stimulus, in both time and anatomical space. In other words, source localization is used to decompose the raw signal based on likely anatomical origin, and this decomposition is then used to estimate each potential source location's response to a particular stimulus variable.

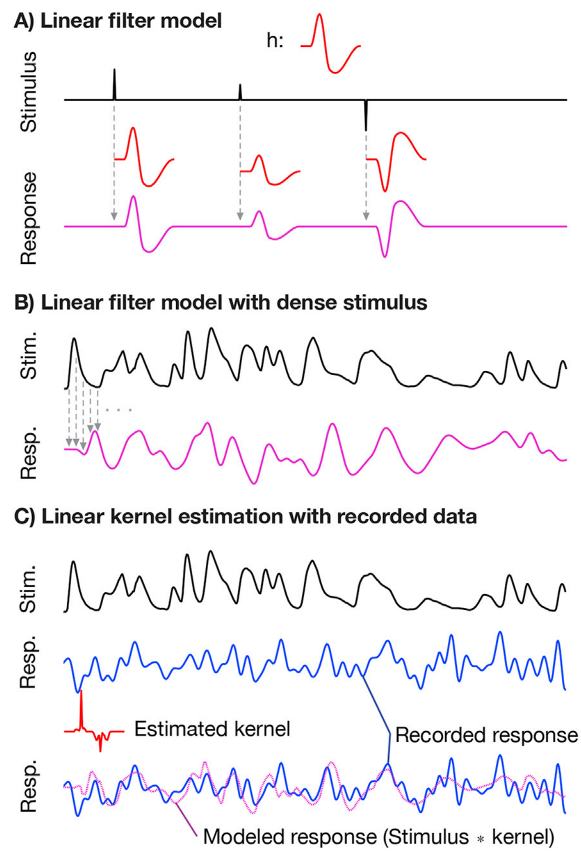

Fig. 1. Linear filter model. The linear filter model $r = h*s$ assumes that the response r is the convolution of the stimulus s with a response function, or kernel h. A) Illustrates the linear filter model for a simple stimulus with three discrete impulses. Since the impulses are spaced far apart relative to the size of the kernel, the shape of the kernel is apparent in the response. B) If the stimulus varies continuously, the convolution leads to a response in which the kernel is not discernible by eye. The response shown is obtained by convolving the stimulus with the same kernel as in A. C) If the stimulus as well as the response are known, different methods exist to estimate a kernel that optimally predicts the response given the stimulus. In the illustration, the simulated response is obtained by convolving the stimulus with the kernel shown under A and adding noise. The kernel is then estimated from the stimulus and the simulated response using boosting (see Methods). The modeled response, i.e. the stimulus convolved with the estimated kernel, can be compared to the actual response to determine the explanatory power of the model.

To test and demonstrate this procedure, we analyzed data from participants listening to segments of a narrated story. We show that 6 min of data per participant is enough to estimate response functions that are reliable across subjects. In order to demonstrate the ability to localize responses in different brain regions, we focused on predictor variables with clearly different predictions for their anatomical localization and temporal response characteristics (see Fig. 2): the response to the acoustic envelope of the speech signal should be associated with at least two strong components around 50 and 100 ms latency, in auditory cortex; previous studies suggest that the latter component is posterior to the former (Ding and Simon, 2012a). Responses associated with word recognition were assessed via lexical frequency, which is known to be one of the strongest predictors of lexical processing in general (see e.g.