ABSTRACT

Title of Dissertation:	TIME-LOCKED CORTICAL PROCESSING OF SPEECH IN COMPLEX ENVIRONMENTS
	Joshua Pranjeevan Kulasingham, Doctor of Philosophy, 2021
Dissertation directed by:	Professor Jonathan Z. Simon, Department of Electrical and Computer Engineering

Our ability to communicate using speech depends on complex, rapid processing mechanisms in the human brain. These cortical processes make it possible for us to easily understand one another even in noisy environments. Measurements of neural activity have found that cortical responses time-lock to the acoustic and linguistic features of speech. Investigating the neural mechanisms that underlie this ability could lead to a better understanding of human cognition, language comprehension, and hearing and speech impairments.

We use Magnetoencephalography (MEG), which non-invasively measures the magnetic fields that arise from neural activity, to further explore these time-locked cortical processes. One method for detecting this activity is the Temporal Response Function (TRF), which models the impulse response of the neural system to continuous stimuli. Prior work has found that TRFs reflect several stages of speech processing in

the cortex. Accordingly, we use TRFs to investigate cortical processing of both low-level acoustic and high-level linguistic features of continuous speech.

First, we find that cortical responses time-lock at high gamma frequencies (~100 Hz) to the acoustic envelope modulations of the low pitch segments of speech. Older and younger listeners show similar high gamma responses, even though slow envelope TRFs show age-related differences. Next, we utilize frequency domain analysis, TRFs and linear decoders to investigate cortical processing of high-level structures such as sentences and equations. We find that the cortical networks involved in arithmetic processing dissociate from those underlying language processing, although both involve several overlapping areas. These processes are more separable when subjects selectively attend to one speaker over another distracting speaker. Finally, we compare both conventional and novel TRF algorithms in terms of their ability to estimate TRF components, which may provide robust measures for analyzing group and task differences in auditory and speech processing. Overall, this work provides insights into several stages of time-locked cortical processing of speech and highlights the use of TRFs for investigating neural responses to continuous speech in complex environments.

TIME-LOCKED CORTICAL PROCESSING OF SPEECH IN COMPLEX ENVIRONMENTS

by

Joshua Pranjeevan Kulasingham

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Professor Jonathan Z. Simon, Chair
Professor Shihab Shamma
Professor Steve Marcus
Associate Professor Behtash Babadi
Associate Professor Ellen Lau

Dedication

To Almighty God, for giving me the freedom and capability to study His works

Acknowledgements

I would like to extend my deepest gratitude to my advisor Prof. Jonathan Simon, for his constant encouragement and guidance. He provided a research environment that was captivating from both an engineering and a scientific perspective. I am thankful not only for his invaluable research supervision, but also for his availability, positivity, support and attentive mentorship throughout all aspects of my PhD career. His passion inspires me to aim high and reminds me to never lose the wonder and delight of 'doing science'.

I am thankful to Prof. Shihab Shamma, Prof. Steve Marcus, Prof. Behtash Babadi and Prof. Ellen Lau for agreeing to serve on my dissertation committee, and for all their valuable feedback. A special thanks to Prof. Behtash Babadi and Prof. Steve Marcus for their insights on signal processing and estimation during our many meetings. Thanks also to Prof. Ellen Lau for introducing me to the fascinating world of cognitive neuroscience and for her guidance on the course project that was the precursor to some of my research work.

I am grateful to all my collaborators who have worked with me on various projects. Thanks to Christian Brodbeck, for his immense support and guidance, for patiently answering all my many questions, and for inspiring me to perform research with greater scientific rigor. Thanks to Prof. Samira Anderson and Prof. Stefanie Kuchinsky for their expert insights and valuable feedback on my work. Thanks to Neha Joshi and Mohsen Rezaeizadeh for their tireless work and ideas when brainstorming, conducting experiments and analyzing data, and for persevering to the end of a long project.

Thanks to Dr. Elisabeth Marsh for providing me with an exciting opportunity to work with stroke patients, and for her patience, insights, and guidance throughout the project. Thanks to Alessandro Presacco and Lien Decruy for all their technical and career guidance and for brightening up the atmosphere at our lab. Thanks to Peng Zan, Dushyanthi Karunathilake, Anuththara Rupasinghe, Proloy Das and Behrad Soleymani for helping me navigate graduate school, and for sharing their knowledge and ideas on my research. Thanks also to Anna Namyst, Ciaran Stone and all my other colleagues who helped with MEG data collection, analysis, and interpretation.

On a personal note, I would like to thank all my friends, especially those from IFC, church and Sri Lanka, who have made the last few years unforgettable. Thanks to my Bible study group, for all the stimulating discussions, spiritual support and enjoyable experiences. Thanks to my uncle Sujeevan, aunt Anuja, and cousin Annika for helping me transition to a new country and for providing a taste of home away from home. Thanks to my parents, Sathy and Rushira, and my sister, Shamika, for their unwavering support, wisdom and constant reminders to keep things in perspective. Finally, I am immeasurably grateful to my wife, Evanjalin, for being my comfort and my anchor, with her love and clear thinking in the midst of every situation.

"Great are the works of the LORD, studied by all who delight in them." – Psalm 111:2

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Our ability to comprehend speech depends on an intricate chain of neural mechanisms that transform the sound waves that enter the ear into meaningful representations in the brain. In addition, speech signals are often mixed with other sounds since we interact with each other in noisy environments. And yet, we are able attend to the relevant speaker with an ease that is not matched by even the most advanced artificial algorithms. Understanding how the brain achieves these tasks, such as the segregation of one speaker among many, or the representation and processing of relevant acoustic and linguistic features, is a vibrant area of research. Investigating these neural mechanisms could further our knowledge on human cognition and lead to improvements in a wide range of applications including the diagnosis and treatment of hearing and speech disabilities, smart hearing aids and automated speech recognition systems.

Extracting information from speech is challenging partly due to the inherent characteristics of the acoustic signals. These signals have rich temporal and spectral properties, and require rapid processing mechanisms (Chi et al., 2005; Rosen, 1992). Remarkably, the human brain is capable of tracking relevant features of acoustic signals in a time-locked manner, faithfully encoding speech with high temporal fidelity (Aiken and Picton, 2008a; Nourski et al., 2009). Magnetoencephalography (MEG), which non-

invasively measures the magnetic fields that arise from the electrical currents within the brain, is well suited to measure such rapid neural processes, due to its fine temporal resolution (Ahissar et al., 2001; Hämäläinen et al., 1993; Luo and Poeppel, 2007). However, the magnetic signals arising from these processes are barely detectable from outside the scalp, and are often contaminated with irrelevant neural activity or environmental noise.

To overcome this challenge, traditional methods rely on averaging the neural responses to hundreds of trials of repeated acoustic stimuli, in order to enhance consistent activity across trials (Picton, 2013; Picton et al., 1974). However, these methods are not suitable for exploring cortical responses to non-repetitive, long duration, continuous speech. Therefore, linear models of neural activity called Temporal Response Functions (TRFs) have been utilized to investigate time-locked neural responses to continuous stimuli without the need for averaging over trials (Ding and Simon, 2014; Lalor and Foxe, 2010). Both TRF and frequency domain analyses have also found neural tracking of linguistic features such as phonemes, words and sentences (Brodbeck et al., 2018a; Ding et al., 2016). TRF analysis has also been used to investigate age related hearing loss (Brodbeck et al., 2018b) and speech segregation during selective attention (Ding and Simon, 2012a). These techniques provide versatile tools to explore several stages of time-locked cortical processing of speech; from the processing of the acoustics, to the processing of semantics and meaning, ultimately leading to speech comprehension.

This dissertation further investigates time-locked neural responses to continuous speech as measured by MEG. We explore time-locked cortical processing of both low-level acoustic features (Chapter 3) and high-level structures like spoken sentences and equations (Chapter 4). TRF analysis often involves estimating and detecting changes in TRF components across groups or tasks to investigate the underlying cortical activity. Hence, we compare several algorithms in terms of their ability to estimate these TRF components (Chapter 5). A more detailed description of this dissertation is given below.

Chapter 2 provides background information on the neural basis, instrumentation and analysis of MEG signals, and reviews neural responses to sound and speech as measured by MEG.

Chapter 3 explores high gamma (70-200 Hz) responses to continuous speech in younger and older adults. MEG responses to continuous speech are typically analyzed in low frequency ranges (~1-10 Hz), for two reasons. 1) The MEG signal to noise ratio reduces with increasing frequency (Hansen et al., 2010) 2) The cortical activity that predominates the MEG signal rarely time-locks to speech at these high rates (Miller et al., 2002). However, recent work has found cortical Frequency Following Responses (FFRs) to repetitive speech syllables in the ~100 Hz range using MEG (Coffey et al., 2017b, 2016). Prior work has shown that the FFR could be useful for diagnosing hearing deficiencies, including age-related hearing loss (Anderson et al., 2012; Kraus et al., 2017a). In this study, we show that MEG is also able to detect high gamma responses to continuous speech, possibly coming from thalamic inputs to auditory

cortex. Using TRF analysis, we find these responses are predominantly driven by the envelope modulation (more than the carrier) of the low pitch segments of the speech stimuli. Interestingly, older and younger listeners have similar high gamma responses, even though FFRs and low frequency cortical TRFs show age-related differences. This work has been published in NeuroImage (Kulasingham et al., 2020).

Chapter 4 investigates high-level cortical processing of spoken arithmetic equations and non-math sentences in a complex two-speaker environment, using frequency domain analysis, TRFs and linear decoders. Cortical responses track hierarchical structures such as phrases and sentences and are enhanced by attention (Ding et al., 2018). Prior studies have found evidence for a dedicated cortical network that underlies numerical processing that is distinct from networks involved in linguistic processing (Amalric and Dehaene, 2019, 2018). In this study, we use selective attention, frequency analysis and TRFs to separate cortical processing of language, arithmetic and acoustics. We find that the spatiotemporal patterns underlying time-locked cortical processing of arithmetic show both similarities and differences from those underlying linguistic processing. The responses to equations and non-math sentences are more clearly separable when selectively attending to one speaker in the two-speaker paradigm. The neural tracking of sentence and equation structures is correlated with behavioral performance on an outlier detection task, suggesting that these responses are linked to comprehension. This work has been published in the Journal of Neuroscience (Kulasingham et al., 2021).

Algorithms for estimating TRFs to continuous speech are investigated in Chapter 5. The neural response to the acoustics of continuous speech typically comprises of well-studied components like the M50 (P1) at ~50 ms and the M100 (N1) at ~100 ms (Ding and Simon, 2012b; Picton, 2013). Investigating differences in auditory processing among individuals, groups, or tasks typically relies on the amplitudes and latencies of these peaks. For example, older subjects have exaggerated M50 and M100 amplitudes (Brodbeck et al., 2018b), and the M100 amplitude is modulated by selective attention (Brodbeck and Simon, 2020; Ding and Simon, 2012a). However, current methods for estimating TRFs often result in ambiguous component waveforms, making it difficult to determine subject-specific component latencies and amplitudes. We investigate both current methods and novel algorithms that utilize prior knowledge of the morphology of these responses, and compare their accuracy in estimating component latencies and amplitudes. Estimating reliable subject-specific TRF components may provide robust measures of differences in neural responses across groups or tasks, leading to improved diagnostic measures and understanding of the cortical processing of continuous speech.

Chapter 2

Background

The brain is composed of billions of neurons, each one rapidly receiving and transmitting signals, resulting in the most complex system of information processing known to man. Measuring this neural activity can be done either by inserting electrodes into neural tissue (invasive methods) or from outside the scalp (non-invasive methods). Neural signals are harder to detect from far away, and non-invasive methods typically result in reduced resolution, accuracy, and precision. Although invasive methods are commonly used in animal research, exploring higher cognitive functions such as speech and language can only be done with human subjects. However, invasive methods involving contact with neural tissue are only feasible for human subject research in rare cases (e.g., during brain surgery on patients with neurological disorders). Therefore, the vast majority of research on speech processing in the human brain is conducted using non-invasive methods such as functional Magnetic Resonance Imaging (fMRI), Electroencephalography (EEG) and Magnetoencephalography (MEG). These methods measure aggregate activity of thousands of neurons, which together form rich patterns of activity across cortical areas.

fMRI indirectly measures neural activity by detecting changes in blood oxygenation in the brain and is one of the most widely used non-invasive methods due to its high spatial resolution. However, changes in blood flow occur at much slower rates than the underlying neural activity, resulting in fMRI having poor temporal resolution (in the order of seconds). Therefore, fMRI is not suited for studying the rapid

mechanisms involved in time-locked speech processing. In contrast, EEG directly measures the electrical currents that arise from neural activity and has very high temporal resolution in the order of milliseconds, at the cost of having low spatial resolution (Lopes da Silva, 2013). The complementary method of MEG measures the magnetic fields elicited by these electrical currents and provides similar temporal resolution and improved spatial resolution (Lopes da Silva, 2013). Therefore, MEG is well suited to investigate the neural mechanisms underlying speech tracking (Ahissar et al., 2001; Luo and Poeppel, 2007). An overview of the neural basis, instrumentation, data collection and typical data analysis techniques for MEG is provided in section 2.1, while section 2.2 provides an overview of time-locked MEG responses that reflect neural processing of sound and speech.

2.1. Magnetoencephalography: Extracting Meaning from Neural Signals

2.1.1. Neural basis and instrumentation of Magnetoencephalography

The fundamental processing unit of the brain is the neuron, which transmits and receives information in the form of electrochemical signals. These signals give rise to magnetic fields, which vary in conjunction with neuronal activity. The magnetic field arising from the current flow between individual neurons is too weak to be measured from a distance. However, several neurons are active at any given moment, and if the electrical fields involved in this activity are synchronized and aligned, the resulting

magnetic fields can be detected even from outside the head. MEG measures these magnetic fields which arise from the aggregate activity of large populations (thousands or even millions) of neurons (Baillet, 2017; Hämäläinen et al., 1993).

Two types of electrical activity in neurons give rise to magnetic fields: fast action potentials (AP) transmitted along the axon, and slower post synaptic potentials (PSP) that arise in the dendrites. Since action potentials are so fast (~1 ms), they are not often synchronized across many neurons. PSPs on the other hand are slow enough (~10 ms) to elicit synchronized magnetic fields across large populations of neurons. These PSPs give rise to Local Field Potentials (LFP) and the corresponding magnetic fields can be detected at a distance using MEG (Hämäläinen et al., 1993).

Although MEG is able to detect currents originating deep in the brain, the sensitivity of MEG to deep sources is very poor, especially when cortical sources are active concurrently (Hansen et al., 2010; Hillebrand and Barnes, 2002). MEG signals also do not capture much information at high frequencies (above 100 Hz), unlike EEG, because the cortical sources that drive MEG rarely synchronize at these frequencies (Miller et al., 2002). However, the MEG signal is much cleaner than EEG, with a higher spatial resolution, since the intervening materials of the cerebrospinal fluid and skull interfere with and distort electrical signals, but are more transparent to magnetic fields (Lopes da Silva, 2013). In Chapter 3, we investigate high frequency time-locked responses to speech, that are traditionally thought to originate in subcortical areas, and find that they also arise from cortical areas.

The MEG system consists of superconducting quantum interference detectors (SQUIDs), that measure the magnetic field and are arranged in a spherical array around the head. Typical MEG systems comprise of around 100 – 300 such sensors placed around the scalp, allowing for good spatial resolution of magnetic fields. These sensors need to be placed inside the insulated 'dewar' and are cooled to superconducting temperatures of around 4 K with liquid Helium. These sensors are sensitive enough to detect the weak magnetic signals produced by neural activity. MEG measurements are performed inside a magnetically shielded room in order to reduce interference from strong magnetic fields and ensure high data quality. Some amount of interference from unwanted magnetic sources is unavoidable, and the next section reviews commonly used techniques to mitigate this problem.

2.1.2. Denoising of MEG signals

The neural signals measured by MEG are typically buried under unwanted magnetic signals from a multitude of sources. The magnetic fields in the brain are orders of magnitude smaller than magnetic signals from other sources like laboratory noise, biological artifacts, or even the earth's magnetic field (Hämäläinen et al., 1993). Sources of noise can be categorized into external noise (eg: magnetic fields arising from laboratory equipment), sensor noise (due to imperfections in the measurement sensors), biological artifacts (signals arising from the participant due to eyeblinks, heartbeat and movements) and background neural activity (neural activity that is not relevant to the current experiment).

Two methods that are commonly used for denoising are Time Shift Principle Component Analysis (TSPCA) (de Cheveigné and Simon, 2007), and Sensor Noise Suppression (SNS) (de Cheveigné and Simon, 2008). TSPCA uses reference sensors to regress out environmental magnetic signals from the MEG signals. The SNS algorithm seeks to eliminate sensor noise by smoothing the measurements of each sensor based on its neighbors. For the work reported in this dissertation, these two methods were used prior to data analysis. However, even small movements of a participant cause magnetic fluctuations far greater than neural signals. In addition, eyeblinks, eye movements and heartbeats are often unavoidable generators of large magnetic signals. Independent Component Analysis (ICA) is a commonly used algorithm for denoising of such artifacts (Barbati et al., 2004; Vigário et al., 1998) and was used in Chapter 4 of this work.

Although subjects are asked to perform a specific task during the MEG recording, a host of other neural processes are active concurrently. Denoising Source Separation (DSS) is an algorithm that seeks to find a spatial filter over sensor space that enhances auditory responses and minimizes contributions from these background neural processes (Cheveigné and Simon, 2008). When an auditory stimulus is repeatedly presented in multiple trials, it evokes an auditory response that is similar across all the presentations. Therefore, DSS finds the spatial filter that maximizes the consistent response across trials and can be used for both denoising as well as dimensionality reduction. In Chapter 5, a subset of DSS components that represent the auditory component of the neural activity were used for some parts of the analysis.

2.1.3. Source localization of MEG responses

The MEG measurement comprises of multiple sensors outside the scalp that measure magnetic fields that arise from neural currents. For simplicity, these magnetic fields are assumed to arise from a fixed number of current dipoles located in the brain. The relationship between the MEG sensor measurements and the current dipoles is given by Maxwell's equations of electromagnetism, and estimating these current dipoles from MEG signals is called source localization (Hansen et al., 2010). This problem is highly underdetermined given that the number of sensors (100-300) is orders of magnitude smaller than the number of current dipoles (>1000). Since several source configurations can give rise to the same magnetic field patterns at the sensors, the source localization is not unique, and additional criteria must be used to select one solution. One of the most popular methods is minimum norm estimation (MNE) (Hämäläinen and Ilmoniemi, 1994). The source localization model can be stated as,

$$\mathbf{Y} = \mathbf{L}\,\mathbf{X} + \mathbf{V} \tag{2.1}$$

Where $\mathbf{Y} \in \mathbb{R}^{M \times T}$ is the measured sensor data over M sensors and T time points, $\mathbf{X} \in \mathbb{R}^{S \times T}$ is the unknown source activity matrix with S sources and T time points that needs to estimated, $\mathbf{V} \in \mathbb{R}^{M \times T}$ is the unknown measurement noise, and $\mathbf{L} \in \mathbb{R}^{M \times S}$ is the lead field matrix that maps the sources onto the sensors. This forward mapping $\mathbf{L}$ can be derived by modelling the brain with the aid of anatomical fMRIs and applying Maxwell's equations to determine the magnetic fields induced by each source.

Therefore, the MNE algorithm estimates the source activity matrix **X** by solving the following minimization problem.

$$\widehat{\mathbf{X}} = argmin \, \|\mathbf{Y} - \mathbf{L}\,\mathbf{X}\|_{\mathbf{C}^{-1}} + \lambda^2 \|\mathbf{X}\|_{\mathbf{R}^{-1}} \tag{2.2}$$

where $\|\mathbf{X}\|_{\mathbf{A}} = \mathbf{X}^{\mathbf{T}}\mathbf{A}\mathbf{X}$, $\mathbf{C} \in \mathbb{R}^{M \times M}$ is the sensor space measurement noise covariance, $\mathbf{R} \in \mathbb{R}^{S \times S}$ is the source covariance, and λ^2 is a regularization parameter. MNE source estimation tends to favor sources close to the surface, and hence modifications such as dynamical Statistical Parametric Mapping (dSPM) (Dale et al., 2000) are used to reweight the source covariance matrix to reduce this bias against deep sources (and has been used in Chapter 3). Given the low spatial resolution of MEG, minimum norm estimation performs reasonably well at estimating distributed cortical activity, and is used for all the source localization analysis done in this dissertation.

2.1.4. The Temporal Response Function (TRF): A Linear Model of MEG Responses

Traditional methods for analyzing MEG responses rely on averaging over multiple trials of repeated stimuli in order to cancel out the background noise (Picton, 2013). Although these methods are suitable for exploring time-locked responses to simple sounds such as tones or speech syllables, they cannot be used to investigate neural processing of continuous speech. Linear models have been proposed to either decode continuous speech stimuli from the neural responses (decoding model), or to find a mapping from features of the speech onto the neural response (encoding model). One

such encoding model is the Temporal Response Function (TRF) given in Eq. 2.3. and shown in Fig. 2.1. (Ding and Simon, 2012a; Lalor and Foxe, 2010).

$$y(t) = \sum_{d} \tau(d)x(t-d) + n(t) \tag{2.3}$$

where $y(t)$ is the measured response at time t, $x(t-d)$ is the time shifted predictor (e.g., acoustic envelope of speech), with a time lag of d, $\tau(d)$ is the TRF value at lag d and $n(t)$ is the residual noise. TRFs model the neural response to speech as a linear filter, with neural processing reflected in the amplitudes and latencies of the lagged components of this filter. This filter is analogous to the evoked response to simple sounds, and can be thought of as the impulse response of the neural system.

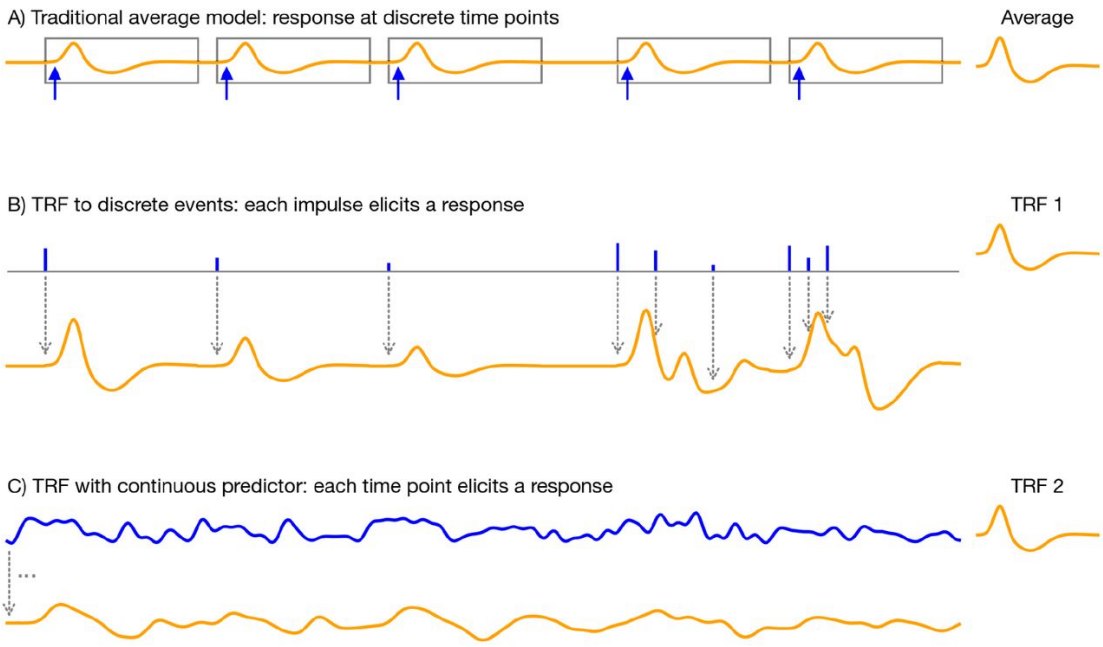

Figure 2.1. TRFs for continuous speech processing. (A) Traditional evoked response model. The evoked response is the average response over many trials. (B) TRF to discrete events. The measured

neural signal is formed by a convolution of the TRF and the stimulus predictor. In this case, the predictor is composed of impulses that represent discrete stimulus events. The size of the impulse determines the magnitude of the response. (C) TRF to a continuous predictor. The measured neural signal is given by the convolution of the TRF and the continuous predictor. *Adapted from Brodbeck et al., 2021b.*

This dissertation focuses only on the TRF encoding model. The TRF can be estimated either for each MEG sensor signal, for the auditory DSS components or for the source localized signals at each neural source. Several algorithms have been used to estimate TRFs including ridge regression (Broderick et al., 2018) and boosting (David et al., 2007). The boosting algorithm was used for estimating TRFs in Chapters 3 and 4. Starting from an all-zero TRF, at each iteration, the boosting algorithm greedily assigns a small discrete increment or decrement to a particular lag in the TRF, that best minimizes the squared error between the predicted and the measured signals. The iteration terminates when the correlation between the predicted and measured signals stops improving. This leads to sparse TRFs that capture only the neural activity that best predicts the response. However, the above algorithms are agnostic to the morphology and structure of neural responses. Chapter 5 investigates both these algorithms as well as novel algorithms that directly estimate TRF components based on prior knowledge of auditory responses.

2.2. Time-Locked Neural Responses to Sound and Speech

2.2.1. The human auditory pathway

From the moment a speech signal enters the ear, a complex chain of processes is set in motion that eventually results in comprehension. The pressure fluctuations of the acoustic waveform travels through the ear canal and vibrates the eardrum (tympanic membrane), which in turn vibrates the middle ear bones (ossicles) (see Fig 2.2A). These vibrations are transferred to the cochlea, where inner hair cells convert these vibrations to electrical signals to be transmitted along the auditory nerve. The electrical signals in the auditory nerve travel through several intermediary structures along the ascending auditory pathway before reaching the primary auditory cortex (see Fig 2.2B). Although the activity of several structures along the ascending auditory pathway can be detected with EEG, cortical responses dominate MEG measurements, since it is not as sensitive to deep structures (Hansen et al., 2010).

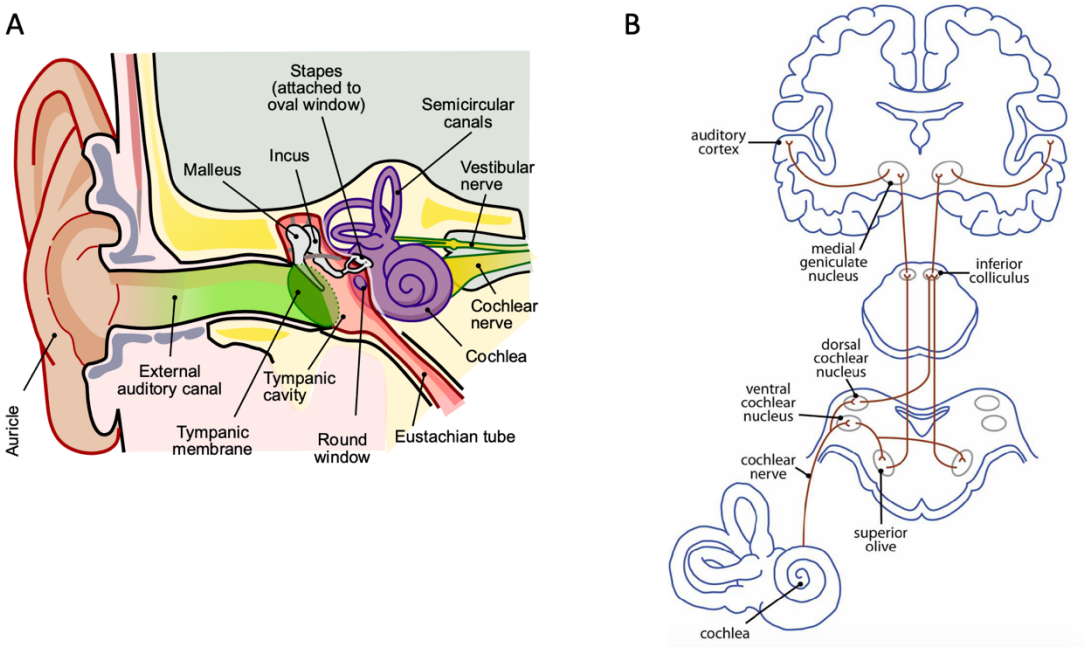

Figure 2.2. The human auditory system. A. Schematic of the human ear. Sound travels through the ear canal and vibrates the tympanic membrane and ossicles. These vibrations are transferred to the cochlea where inner hair cells convert them to electrical signals to be transmitted along the auditory nerve. *Adapted from Chittka and Brockmann, 2005*. **B.** Schematic of the ascending auditory pathway. Auditory information travels from the cochlea, through several intermediate subcortical processing structures, until finally arriving at the auditory cortex. *Adapted from Butler and Lomber, 2013*

2.2.2. MEG for auditory responses

There has been a long history of research on the auditory system using EEG and MEG. Traditionally, the measured signal is averaged over many trials with the same auditory stimulus in order to cancel out the background noise (Picton et al., 1974). This method can detect evoked responses that are time-locked to the onset of the stimulus.

The activity of each neural structure along the ascending pathway leads to measurable evoked responses which are broadly categorized into canonical early (under 10 ms), middle (10 – 50 ms) and late (above 50 ms) components (Picton, 2013). The late components are most relevant for MEG measurements of speech processing, since these arise from primary and secondary auditory cortex, as well as from cortical areas reflecting higher order processing. Canonical late auditory components comprise of a positive peak at 50 ms termed the P1 (or M50 for MEG measurements), a negative peak at 100 ms termed N1 (or M100) and a positive peak around 200 ms termed the P2 (or M200). These peaks are thought to reflect different stages of neural processing and have been widely studied to investigate age-related hearing loss (Tremblay et al., 2003), auditory disorders (Picton, 2013), and attentional modulation (Näätänen, 1990).

The auditory system also time-locks to the fundamental frequency of an acoustic stimulus, resulting in the Frequency Following Response (FFR) which is typically studied using EEG (see Fig 2.3). The FFR could be helpful in understanding age related hearing loss and other hearing impairments (Kraus et al., 2017a). The FFR is a very fast response in the range of 100 – 1000 Hz and predominantly arises from subcortical sources. However, recent studies using MEG and EEG have shown that cortical areas contribute to the ~100 Hz FFR to repeated speech syllables (Coffey et al., 2016, 2017b). In Chapter 3, we investigate time-locked MEG responses to continuous speech that are in the FFR range and find that these responses originate from the cortex and are consistent across younger and older listeners.

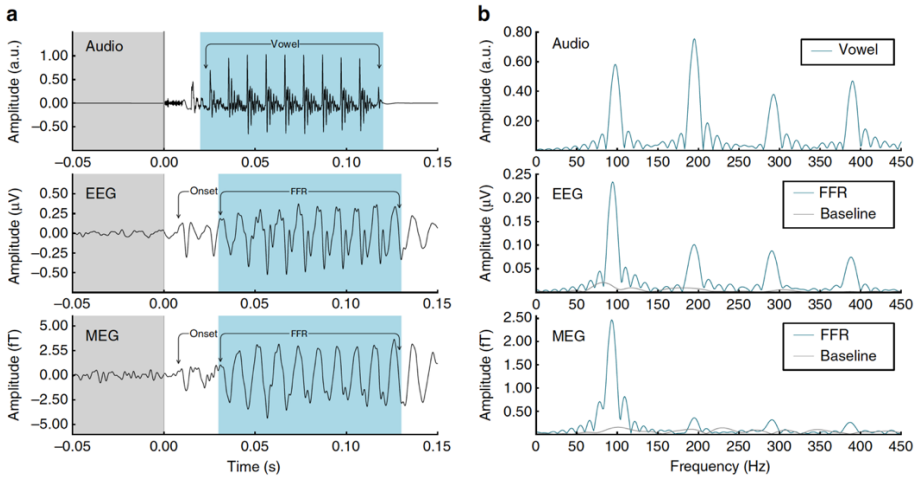

Fig 2.3. Frequency following responses. **a.** The stimulus waveform (speech syllable /da/) is shown along with the EEG and MEG responses averaged over several hundreds of repetitions. The FFR is clearly visible in both EEG and MEG for the duration of the vowel. **b.** The frequency spectrum shows that the EEG and MEG responses time-lock predominantly to the fundamental frequency of the audio signal. *Adapted from Coffey et al., 2016.*

2.2.3. MEG responses to continuous speech using TRFs

When the stimulus is continuous speech, averaging over multiple trials is no longer feasible, and instead TRFs have been used (Ding and Simon, 2012b; Lalor and Foxe, 2010). The TRF waveform shows consistent M50, M100 and M200 peaks similar to the evoked response. These TRFs are typically estimated for the auditory envelopes of the speech stimuli. However, TRF analysis can also be used to investigate processing of higher order features of speech including words and semantics (Brodbeck et al., 2018a; Broderick et al., 2018). In Chapter 4, we use TRFs to estimate neural responses to sentences and equations and find sustained responses that vary over several cortical areas.

2.2.4. TRFs to speech in complex environments

The cocktail party paradigm has been commonly used for investigating cortical speech processing in complex environments (Cherry, 1953; Middlebrooks et al., 2017). Subjects are asked to attend to one speaker in the presence of one or more background speakers, simulating the common experience of listening to one person in the midst of a crowd, such as in a cocktail party. Since the speech waveforms of both speakers are mixed into one acoustic signal, segregating the relevant speech stream is quite a challenging task. By using the foreground and background stimuli as predictors, TRFs to both the attended and unattended speech stream can be estimated. The early M50 peak is present for both foreground and background TRFs, indicating that it reflects pre-attentive auditory processing. However, the M100 peak shows strong attentional modulation (Ding and Simon, 2012b; Zion Golumbic et al., 2013), suggesting that the speech streams are at least partially segregated around 100 ms after the stimulus enters the periphery.

Studies utilizing such TRF models typically contrast the amplitudes and latencies of TRF components across groups or tasks. Hence robust estimates of both group effects as well as individual TRFs are essential. However, single subject TRFs are often very noisy and may not have clear component peaks. In Chapter 5, we investigate TRF algorithms and compare their ability to estimate the amplitudes and latencies of well-known TRF components. Such algorithms could pave the way toward further understanding cortical processing of speech and improvements in the treatment and diagnosis of hearing and speech impairments.

Chapter 3

High Gamma Cortical Processing of Continuous Speech in Younger and Older Listeners

This work has been published as

Kulasingham, J.P., Brodbeck, C., Presacco, A., Kuchinsky, S.E., Anderson, S., Simon, J.Z., 2020. High gamma cortical processing of continuous speech in younger and older listeners. NeuroImage 222, 117291.

https://doi.org/10.1016/j.neuroimage.2020.117291

3.1. Abstract

Neural processing along the ascending auditory pathway is often associated with a progressive reduction in characteristic processing rates. For instance, the well-known frequency-following response (FFR) of the auditory midbrain, as measured with electroencephalography (EEG), is dominated by frequencies from ~100 Hz to several hundred Hz, phase-locking to the acoustic stimulus at those frequencies. In contrast, cortical responses, whether measured by EEG or magnetoencephalography (MEG), are typically characterized by frequencies of a few Hz to a few tens of Hz, time-locking to acoustic envelope features. In this study we investigated a crossover case, cortically generated responses time-locked to continuous speech features at FFR-like rates. Using MEG, we analyzed responses in the high gamma range of 70–200 Hz to continuous

speech using neural source-localized reverse correlation and the corresponding temporal response functions (TRFs). Continuous speech stimuli were presented to 40 subjects (17 younger, 23 older adults) with clinically normal hearing and their MEG responses were analyzed in the 70–200 Hz band. Consistent with the relative insensitivity of MEG to many subcortical structures, the spatiotemporal profile of these response components indicated a cortical origin with ~40 ms peak latency and a right hemisphere bias. TRF analysis was performed using two separate aspects of the speech stimuli: a) the 70–200 Hz carrier of the speech, and b) the 70–200 Hz temporal modulations in the spectral envelope of the speech stimulus. The response was dominantly driven by the envelope modulation, with a much weaker contribution from the carrier. Age-related differences were also analyzed to investigate a reversal previously seen along the ascending auditory pathway, whereby older listeners show weaker midbrain FFR responses than younger listeners, but, paradoxically, have stronger cortical low frequency responses. In contrast to both these earlier results, this study did not find clear age-related differences in high gamma cortical responses to continuous speech. Cortical responses at FFR-like frequencies shared some properties with midbrain responses at the same frequencies and with cortical responses at much lower frequencies.

3.2. Introduction

The human auditory system time-locks to acoustic features of complex sounds, such as speech, as it extracts and encodes relevant information. The characteristic frequency of such time-locked activity is generally thought to decrease along the ascending auditory pathway. For example, subcortical activity at ~100 Hz and above may directly encode the temporal pitch information of voiced speech (Forte et al., 2017; Krishnan et al., 2004), while cortical activity below ~10 Hz, which time-locks to the slowly varying envelope of speech, also time-locks to higher level features of language such as phoneme and word boundaries (Brodbeck et al., 2018a). Prior research has also found differences in both subcortical and cortical processing for older and younger listeners (Anderson et al., 2012; Presacco et al., 2016a, 2016b), which suggest age-related auditory temporal processing deficits. These effects have been investigated in human subjects using the complementary non-invasive neural recording techniques of electroencephalography (EEG) and magnetoencephalography (MEG).

The well-known frequency following response (FFR) is one such phase-locked response (Kraus et al., 2017b), most commonly measured using EEG, and is believed to originate predominantly from the auditory midbrain (Bidelman, 2015; Smith et al., 1975). The FFR measures the phase-locked response to the fast (~100 Hz and above), steady state oscillation of a stimulus, such as a repeated speech syllable. The FFR provides insight into the peripheral representation of speech and is a useful tool for investigating temporal processing deficits (Basu et al., 2010; Hornickel et al., 2012; Kraus et al., 2017b). In addition, the FFR may be used to investigate the robustness of

speech representations in noise or a dual stream paradigm (Yellamsetty and Bidelman, 2019). The FFR is believed to detect the integrated activity of several nonlinear processing stages along the auditory pathway, and hence various nonlinear features of the stimulus can contribute to the FFR (Lerud et al., 2014). Some studies compare and contrast FFRs obtained by averaging or by subtracting responses to stimuli of opposite polarity in order to tease apart these contributions to some extent (Aiken and Picton, 2008; Hornickel et al., 2012).

The neural origins of the FFR have historically been thought to be mainly subcortical areas such as the inferior colliculus (Smith et al., 1975). But recent studies with MEG and EEG have shown that the FFR at ~100 Hz is not purely generated by subcortical areas, but has contributions from the auditory cortex as well (Bidelman, 2018; Coffey et al., 2017b, 2017a, 2016; Hartmann and Weisz, 2019; Puschmann et al., 2019). Some studies have shown that this cortical contribution is stronger in the right hemisphere (Coffey et al., 2016; Hartmann and Weisz, 2019). The dominantly cortical role in the MEG FFR follows from the reduced sensitivity of gradiometer-based MEG to deep structures such as the auditory midbrain (Baillet, 2017).

However, the repeated speech syllables commonly used to generate the FFR cannot capture the complexities of natural continuous speech. To understand how the brain represents speech in naturalistic environments, cortical low frequency (below ~10 Hz) responses to continuous speech have been widely studied (Peelle et al., 2013). The MEG and EEG response to continuous speech can be represented using Temporal Response Functions (TRFs) (Ding and Simon, 2012b; Lalor et al., 2009) which are

linear estimates of time-locked responses to time varying features of the auditory stimulus. The conventional low-frequency TRF time-locks to the slow (below ~10 Hz) envelope of continuous speech, though the spectrotemporal fine structure of speech can also modulate these cortical low frequency responses (Ding et al., 2014; Ding and Simon, 2012b).

Recently, short latency subcortical EEG responses to continuous speech have been found using TRF analysis (Maddox and Lee, 2018), demonstrating that it is possible to detect fast midbrain responses to continuous speech. Early latency responses that phase lock to the fundamental frequency of speech have also been found to be modulated by attention (Forte et al., 2017). One study has also found cortical high gamma MEG responses to speech stimuli, with latencies near 30 ms, that are time-locked to the ~100 Hz temporal modulation in the envelope of the speech spectrum (up to 2 kHz) (Hertrich et al., 2012). Whether auditory cortex time-locks in the high gamma range to the carrier as well as to the envelope modulation of continuous speech remains unclear.

Further complicating our understanding of the contributions of subcortical and cortical sources to the MEG response is the impact of age-related changes in the auditory pathway (Peelle and Wingfield, 2016). The temporal processing of speech can degrade with age, especially in noisy conditions (Gordon-Salant et al., 2006; He et al., 2008; Hopkins and Moore, 2011). Age-related differences have been found in both the EEG FFR and the MEG low frequency TRF to speech. Older adults have weaker, delayed FFRs with lower phase coherence when compared with younger adults (Anderson et al., 2012; Presacco et al., 2015; Zan et al., 2019). Possible causes include

age-related inhibition-excitation imbalance (Caspary et al., 2008) resulting in a loss of temporal precision (Anderson et al., 2012). In a surprising reversal, older adults' cortex exhibits *exaggerated* low frequency responses (Bidelman et al., 2014; Brodbeck et al., 2018b), even to the point of allowing better stimulus reconstruction via these low frequency cortical responses than in younger adults (Presacco et al., 2016a, 2016b). Several possible explanations, not necessarily exclusive, have been advanced to account for this surprising result, including decrease in inhibition, recruitment of additional brain regions and central compensatory mechanisms (Chambers et al., 2016; Peelle et al., 2010). The fact that fast midbrain responses are reduced with age while slow cortical responses are enhanced might indeed be due to anatomical and physiological differences between midbrain and cortex, but a fair comparison is complicated by the fact that the responses occur at vastly different frequencies. Hence it is entirely unknown whether high gamma cortical responses would show age-related reduction or enhancement.

In this study, we investigated high gamma cortical responses to continuous natural speech using MEG. Unfortunately, MEG responses are known to have relatively poor signal-to-noise ratio (SNR) and decreased power at high gamma frequencies because the cortical sources that dominate MEG responses rarely phase lock in this range at a population level (Lu et al., 2001). In addition, environmental noise and artifacts such as muscular movement can obscure the signal at these higher frequencies (Muthukumaraswamy, 2013). Hence detecting high gamma responses using MEG may require averaging over many trials (as for the FFR), or much longer speech stimuli, to

boost SNR. Similarly, detecting subcortical responses to speech may also require high SNR or longer speech stimuli (Maddox and Lee, 2018).

This work investigates whether such high gamma responses can be detected using MEG with a simple experimental paradigm of short duration. MEG recordings of younger and older subjects listening to only six minutes of continuous speech (narration by a male speaker) were investigated using TRF analysis, and such high gamma time-locked responses are indeed found to be present. Just as the low frequency TRF may be compared to a low frequency evoked response, the high gamma TRF may be compared to the FFR in that they both reflect time-locked activity at the stimulus frequency. 70–200 Hz was chosen as the high gamma range because 70 Hz is near the lower end of typical male voice pitch (and well above the 60 Hz of line noise) while 200 Hz is far above most known auditory responses measured by MEG. In addition, source localization was performed to investigate the cortical and subcortical contributions to these high gamma MEG responses. Only six minutes, as opposed to, e.g., 30 minutes, were chosen for the stimulus duration as being typical for an auditory speech experiment that employs multiple stimulus conditions (e.g., several levels of speech in noise). TRF analysis can then be used to investigate time-locked neural processing of a wide variety of stimulus features, from acoustics to semantics (Brodbeck et al., 2018a) simultaneously in the same, short experimental paradigm.

We focused on the following specific research questions. Firstly, are 70–200 Hz MEG responses to continuous speech time-locked to the carrier or to the envelope modulation of the speech spectrum? Unlike FFR analysis, TRF analysis is able to

explicitly and simultaneously capture distinct response contributions arising from different stimulus features, in this case from envelope modulation and the carrier, allowing direct comparison of the separate contributions of these features to the response. Secondly, are there any age-related differences in these responses, and if so, do they show age-related decrease, like the EEG FFR, or the opposite, like the cortical low frequency TRFs? Additionally, we investigated if these responses were right lateralized as found in the MEG FFR (Coffey et al., 2016). Such right lateralization would also agree with studies showing right hemispheric dominance for pitch processing in core auditory cortex (Hyde et al., 2008). Finally, we investigated if the responses were influenced by the instantaneous pitch of the speech stimulus.

3.3. Methods

3.3.1. Experiment dataset

The experimental dataset used for this study has been previously described in detail by Presacco et al. (2016a, 2016b), but is here supplemented with eight additional older adults with clinically normal hearing (dataset available online (Kulasingham, 2019a)). The combined dataset consisted of MEG responses recorded from 17 younger adults (age 18–27, mean 22.3, 3 male) and 23 older adults (age 61–78, mean 67.2, 8 male), with clinically normal hearing, while they listened to 60 second portions of an audiobook recording of "The Legend of Sleepy Hollow" by Washington Irving (https://librivox.org/the-legend-of-sleepy-hollow-by-washington-irving). All

participants gave informed consent and were paid for their time. Experimental procedures were reviewed and approved by the Institutional Review Board of the University of Maryland. The audio was delivered diotically through 50 Ω sound tubing (E-A-RTONE 3A) attached to E-A-RLINK foam earphones inserted into the ear canal at ~70 dB sound pressure level via a sound system with flat transfer function from 40 to 3000 Hz. The conditions analyzed in this study consist of two passages of 60 seconds duration presented in quiet (i.e., solo speaker), each of which was repeated three times, for a total of six minutes of MEG data per subject. Subjects were asked beforehand to silently count the number of occurrences of a particular word and report it to the experimenter at the conclusion of each trial, in order to encourage attention to the auditory stimuli. Handedness of the participants was assessed with the Edinburgh handedness scale (Oldfield, 1971), which can range from –1 (complete left-dominance) to 1 (complete right-dominance). To exclude lateralization bias due to handedness, all analyses were performed again excluding the 9 subjects scoring below 0.5. The only qualitative change in the results was a loss of right hemispheric dominance in younger subjects (discussed below).

3.3.2. MEG data collection and preprocessing

MEG data was recorded from a 157 axial gradiometer whole head KIT MEG system while subjects were resting in the supine position in a magnetically shielded room. The data was recorded at a sampling rate of 1 kHz with an online 200 Hz low pass filter with a wide transition band above 200 Hz, and a 60 Hz notch filter. Data was

preprocessed in MATLAB by first automatically excluding saturating channels and then applying time-shift principal component analysis (de Cheveigné and Simon, 2007) to remove external noise, and sensor noise suppression (de Cheveigné and Simon, 2008) to suppress channel artifacts. On average, two MEG channels were excluded during these stages. All subsequent analyses were performed in mne-python 0.17.0 (Gramfort, 2013; Alexandre Gramfort et al., 2014) and eelbrain 0.30 (Brodbeck et al., 2019); code available online (Kulasingham, 2019b). The MEG data was filtered in the band 70–200 Hz (high gamma band) using an FIR filter described below, and six 60 second epochs during which the stimulus was presented were extracted for analysis. The band 70–200 Hz was chosen since the pitch of the male speaker typically falls in this range, and 200 Hz is above most known auditory cortical responses. The data was resampled to 500 Hz for all further analysis.

3.3.3. Stimulus representation

As discussed above, prior work on the FFR has shown that time-locked neural responses are sensitive to both the carrier and the envelope of an auditory stimulus. Similarly, time-locked responses to speech in the high gamma range may be driven either by the high gamma carrier, or by high gamma modulation in the envelope of even higher frequencies. Accordingly, two distinct representations of the speech stimulus were used as predictors for the TRF model (see Fig. 3.1). For the former case, the carrier predictor was constructed by resampling the speech waveform to 1 kHz (using the mne-python function ‘resample’) and bandpass filtering from 70–200 Hz

using the same filter as above. This carrier predictor captures the high gamma rate modulation in the speech waveform itself. For the latter case, the envelope modulation predictor was constructed from the high gamma modulation in the envelope of the highpassed stimulus waveform (envelopes are only well-defined when they modulate carriers of higher frequencies than those of the modulations themselves; Rosen, 1992). Specifically, first the speech was transformed into an auditory spectrogram representation by computing the acoustic energy in the speech waveform for each frequency bin in the range 300–4000 Hz at millisecond resolution using a model of the auditory periphery (Yang et al., 1992). The range 300–4000 Hz was chosen in order to have a clear separation between the upper end of the high gamma range (200 Hz) and because the auditory stimulus was presented through air tubes which attenuate frequencies above 4000 Hz. This auditory spectrogram is a 2-dimensional matrix representation of the acoustic envelope over time for different frequency bins. Each frequency bin component of this spectrogram was then filtered using the same 70–200 Hz bandpass filter as above, producing a 70–200 Hz band limited envelope for each bin. Finally, the resulting 2-dimensional matrix was averaged across frequency bins to provide a single signal, resulting in the envelope modulation predictor. Thus, this predictor captures the 70–200 Hz temporal modulation in the 300–4000 Hz envelope of the speech waveform. These two predictors were resampled to 500 Hz and used for all further TRF analysis. Even though the two predictors are correlated at $r = -0.42$, the TRF analysis is able to separate the neural response to each of them (negative

correlations are common for a carrier and the corresponding non-linearly related envelope of another frequency band with different cochlear delays).

The 70–200 Hz bandpass filter was formed using the default FIR filter in mne-python with an upper and lower transition bandwidth of 5 Hz, at 1 kHz sampling frequency, but applied twice in a forward fashion to the data. This resulted in a combined filter of length 1322 with a phase delay of 660 ms. Other bandpass filters were also employed as alternatives, including IIR minimum-phase-delay Bessel filters (results not shown); no results depended critically on the filters used.

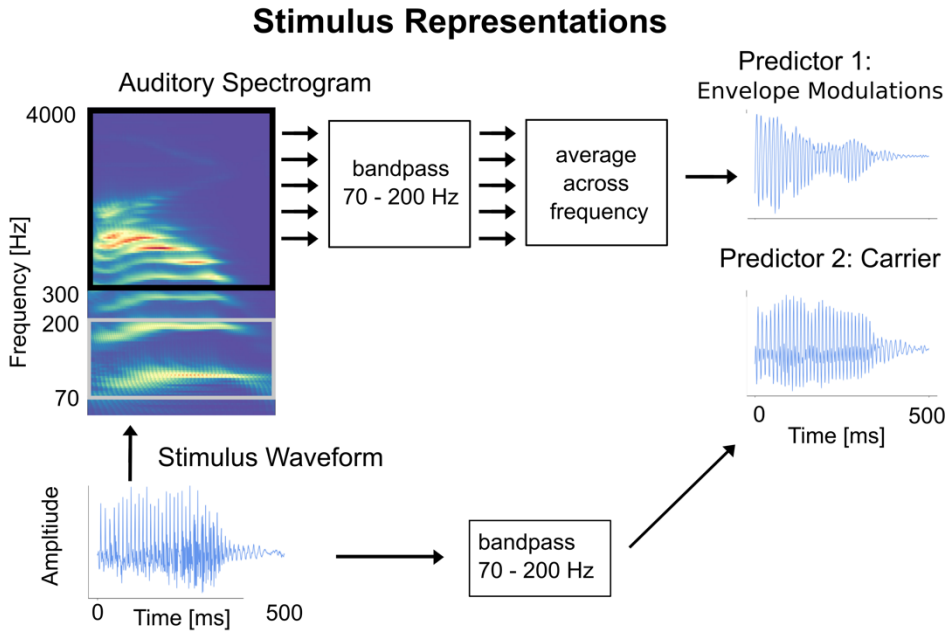

Figure 3.1. Stimulus representations. The stimulus waveform for a representative 500 ms speech segment is shown along with its auditory spectrogram and the two predictors: carrier and envelope modulation. The predictors are correlated (Pearson's r = –0.42) but have noticeably distinct waveforms.

3.3.4. Neural source localization

Before each MEG recording, the head shape of each subject was digitized using a Polhemus 3SPACE FASTRAK system, after which five marker coils were attached. The marker coil locations were measured while the subject's head was positioned in the MEG scanner before and after the experiment, in order to determine the position of the head with respect to the MEG sensors. Source localization was performed using the mne-python software package. The marker coil locations and the digitized head shape were used to coregister the template Freesurfer 'fsaverage' brain (Fischl, 2012) using rotation, translation and uniform scaling. A volume source space was formed by dividing the brain volume into a grid of 7 mm sized voxels. This source space was used to compute an inverse operator using minimum norm estimation (MNE) (Alexandre Gramfort et al., 2014) and dynamical statistical parametric mapping (dSPM) (Dale et al., 2000) with a depth weighting parameter of 0.8, and a noise covariance matrix estimated from empty room data. This method results in a 3-dimensional current dipole vector with magnitude and direction at each voxel. The Freesurfer 'aparc+aseg' parcellation was used to define cortical and subcortical regions of interest (ROIs). The cortical ROI consisted of voxels in the gray and white matter of the brain that were closest to the temporal lobe Freesurfer 'aparc' parcellations ('aparc' labels: 'transversetemporal', 'superiortemporal', 'inferiortemporal', 'bankssts'). A few additional voxels surrounding auditory cortex (within 20 mm) were included in the ROI solely to ensure that the source localized responses not be misleadingly focal (distributed source localization with MNE has a large spatial spread). The subcortical

ROI was selected to consist of voxels that were in the Freesurfer 'aseg' 'Brain-Stem' segmentation. All brain plots show the maximum intensity projection of the voxels onto a 2-dimensional plane, with an overlaid 'fsaverage' brain schematic (implemented in eelbrain). Minimum norm estimation in volume source space may lead to spatial leakage from the true neural source to neighboring voxels. In order to characterize this artifactual spatial leakage, a single current dipole in Heschl's gyrus was simulated, projected into sensor space, and then projected into volume source space (see Appendix). Additionally, a separate cortical surface source space model was also used; results obtained using this method were not qualitatively different than those of the volume space model (see Appendix).

3.3.5. Temporal response functions

The simplest linear model used to estimate the TRF is given by

$$y_t = \sum_d (\tau_d x_{t-d}) + n_t \quad (1)$$

where y_t is the response at a neural source for time t, x_{t-d} is the time shifted predictor with a time lag of d, τ_d is the TRF value at lag d and n_t is the residual noise. The TRF is the set of time-dependent weights, of a linear combination of current and past samples of the predictor, that best predicts the current neural response at that neural source (Lalor et al., 2009). Hence the TRF can also be interpreted as the average time-locked response to a predictor impulse. In this investigation, a TRF model with two predictors, envelope modulation and carrier, was used.

$$y_t = \sum_d (\tau_{e,d} e_{t-d} + \tau_{c,d} c_{t-d}) + n_t \quad (2)$$

Where e_{t-d} is the delayed envelope modulation predictor and $\tau_{e,d}$ the corresponding envelope modulation TRF, $c_{t-d}$is the delayed carrier predictor and $\tau_{c,d}$ the corresponding carrier TRF. In this model, the two predictors compete against each other to explain response variance, which results in larger TRFs for the predictor that contributes more to the neural response. The model parameters were estimated jointly, such that the model is not affected by the ordering of the predictors. TRF estimation, for lags from –40 to 200 ms, was performed with the boosting algorithm and early stopping based on cross validation (David et al., 2007) as implemented in eelbrain. The boosting algorithm may result in overly sparse TRFs, and hence an overlapping basis of 4 ms Hamming windows (with 1 ms spacing) was used in order to allow smoothly varying responses; altering the Hamming window duration did not substantively affect the results. For the volume source space, the neural response at each voxel is a 3-dimensional current vector. Accordingly, for each voxel, a TRF vector was computed using the boosting algorithm and was used to predict the neural response vector. For each voxel, the prediction accuracy was assessed through the average dot product between the normalized predicted and true response, which varies between -1 and 1 in analogy to the Pearson correlation coefficient.

3.3.6. Pitch analysis

Prior studies have suggested that neural time-locking at high gamma rates may reflect processing of pitch related features of the speech (Smith et al., 1978). In order to investigate the extent to which the response oscillations were influenced by the pitch frequency of the speech stimulus, a simple pitch analysis was performed as follows. The pitch of the speech signal was extracted using Praat (Boersma, 1993; Boersma and Weenick, 2018) in sliding 40 ms windows and used to mark times when the pitch was above or below the median pitch value (98.11 Hz). This algorithm is a better approximation of the percept of pitch than simply dividing the stimulus based on its frequency content, thus allowing subsequent analysis to be done on a neurally relevant feature of the stimulus. Two new 'high pitch' predictors were formed based on the previous two predictors (envelope modulation and carrier) by zeroing out the times when the pitch was below the median. Similarly, 'low pitch' predictors were formed by zeroing out times when the pitch was above the median. Time windows without a stable pitch estimate were set to zero in all predictors. Hence four new predictors were created: high pitch envelope modulation, low pitch envelope modulation, high pitch carrier and low pitch carrier. All 4 predictors were used simultaneously in a competing TRF model analogous to that in Eq. 3.2.

3.2.7. Statistical tests

Statistical tests were performed across subjects by comparing the TRF model to a noise model. The predictor was circularly shifted in time and TRFs were estimated using this time-shifted predictors as noise models (Brodbeck et al., 2018a, 2018b). This preserves the local temporal structure of the predictor while removing the temporal relationship between the predictor and the response. Circular shifts of duration 15, 30 and 45 seconds were used to form three noise models. For each voxel, the prediction accuracies of the true model were compared to the average prediction accuracies of the three noise models as a measure of model fit. Since all the predictors in the model are fit jointly, this results in one joint prediction accuracy for all the predictors for each voxel.

To account for variability in neural source locations due to mapping the responses of individual subjects onto the 'fsaverage' brain, these coefficients were spatially smoothed using a Gaussian window with 5 mm standard deviation. Nonparametric permutation tests (Nichols and Holmes, 2002) and Threshold Free Cluster Enhancement (TFCE) (Smith and Nichols, 2009) were used to control for multiple comparisons. This method, as outlined in full in Brodbeck et al., 2018c, 2018a, is implemented in eelbrain, and is briefly recounted here. Firstly, a paired sample *t*-value was evaluated for each neural source, across subjects, from the difference of the prediction accuracies of the true model and the average of the three noise models after rescaling using Fisher's *z*-transform. Then the TFCE algorithm was applied to those *t*-values, which enhanced continuous clusters of large values, based on the assumption

that significant neural activity would have a larger spatial spread than spurious noise peaks. This procedure was repeated 10,000 times with random permutations of the data where the labels of the condition were flipped on a randomly selected subset of the subjects. A distribution of TFCE values was formed using the maximum TFCE value of each permutation to correct for multiple comparisons across the brain volume. Any value of the original TFCE map that exceeded the 95th percentile of the distribution was considered as significant at the 5% significance level. This corresponds to a one-tailed test of whether the true model increases the prediction accuracy over the noise model. In cases where both sides of the comparison are important, corresponding two-tailed tests were used (as explained below, for e.g., left vs. right, younger vs. older, envelope vs. carrier). In all subsequent results, the maximum or minimum t-value across voxels is reported as t_{max} or t_{min} respectively.

The TRF itself was also tested for significance against the noise model in a similar manner. In the volume source space, a TRF that consists of a 3-dimensional vector which varies with time was estimated for each voxel, representing the estimated current dipole amplitude and direction at that voxel. The amplitudes of these TRF vectors for the true model and the average noise model were used for significance testing. The TRF amplitudes were spatially smoothed using the same Gaussian window before performing the tests. A one-tailed test was done with paired sample *t*-values and TFCE, and the procedure is identical to that outlined previously, with the added dimension of time (Brodbeck et al., 2018a).

Lateralization tests were performed to check for hemispheric asymmetry. The volume source space estimates in the cortical ROI were separated into left and right hemispheres and, as above, the prediction accuracies were spatially smoothed with the same Gaussian window. The prediction accuracies of the average noise model were subtracted from that of the true model and paired sample *t*-values with TFCE in a two-tailed test were used to test for significant differences between each of the corresponding left and right voxels.

Age-related differences were assessed between the younger and older groups. The difference of prediction accuracies between the true TRF model and the average of the noise TRF models were used to form independent sample *t*-values for each source across age groups after which a two-tailed test was performed with TFCE. Significant differences in lateralization across age groups were assessed by subtracting the prediction accuracies of the left hemisphere from the right hemisphere and then conducting independent samples tests across age groups as described above. The peak latency of the TRFs was also tested for significant differences across age groups. The latency of the maximum value of the ℓ_2 norm of the TRF vectors in the time range of significant responses (20–70 ms) was used to test for peak latency differences across age groups using a two-tailed test with independent sample *t*-values and TFCE.

To further investigate differences by age across both low frequency and high frequency (i.e., high gamma) responses, two additional models were analyzed; a low frequency (1–10 Hz) TRF and a high frequency TRF with the same parameters as the above models, but using cortical surface source space. An ANOVA was performed on

the prediction accuracies of these two models with factors TRF frequency (high or low) and age (young or old) (detailed methods and results in Appendix section 3.7).

3.4. Results

3.4.1. Cortical origins of high gamma responses to continuous speech

Per-voxel TRFs in volume source space were estimated in the high gamma range for the two ROIs: the temporal lobes, and the brainstem (plus its surrounding volume). The prediction accuracies of the competing stimulus model described above for high-gamma responses (mean = 0.021, std = 0.003) were much smaller (factor of 3) than those resulting from low frequency cortical TRFs (Brodbeck et al., 2018a), indicating that these responses are weaker than slow cortical responses. This is not surprising, as the spectral power of the MEG response decays with frequency. Noise floor models, used to test for significant responses, generated corresponding noise model prediction accuracies (mean = 0.018, std = 0.001). For each voxel, a one-tailed test with paired sample *t*-values and TFCE (to account for multiple comparisons) was used to test for significant increases in the prediction accuracies of the true model against the noise model across subjects. A large portion of the voxels showed a significant increase in prediction accuracy (younger subjects $t_{max} = 6.19$, $p < 0.001$; older subjects $t_{max} = 5.66$, $p < 0.001$; see Fig. 3.2A). The disproportionate extent of this result is not unexpected, however, due to the large spatial spread of MNE volume source space estimates. The prediction accuracy over the noise model for voxels in Heschl's gyrus was significantly

larger than that within the subcortical ROI for both age groups (two-tailed paired sample *t*-test; younger subjects $t = 3.67$, $p = 0.002$; older subjects $t = 2.65$, $p = 0.015$; difference across age not significant). Although some voxels in the subcortical ROI are significant, this can be ascribed to artifactual leakage arising from the source localization algorithm (see simulation in Appendix).

Lateralization differences were tested using the prediction accuracy at each voxel. The prediction accuracy of the average noise model was subtracted from that of the true model and a two-tailed test with paired sample *t*-values and TFCE was performed for significant differences in the left and right hemispheres. The tests revealed significantly higher prediction accuracies for younger subjects in the right hemisphere than in the left ($t_{max} = 3.81$, $p = 0.035$), but only for a few voxels (1.6%) in the temporal lobe close to auditory areas (see Fig. 3.2B). No significant differences in lateralization were seen for older subjects ($t_{max} = 3.41$, $t_{min} = –1.52$, $p > 0.09$), nor was lateralization significantly different across age groups (independent samples test; $t_{max} = 1.93$, $t_{min} = –2.28$, $p > 0.88$). When the analysis was constrained to only right-handed subjects (13 younger, 18 older; see Methods for details), the only resulting change was that no voxels were significantly right lateralized in either age group.

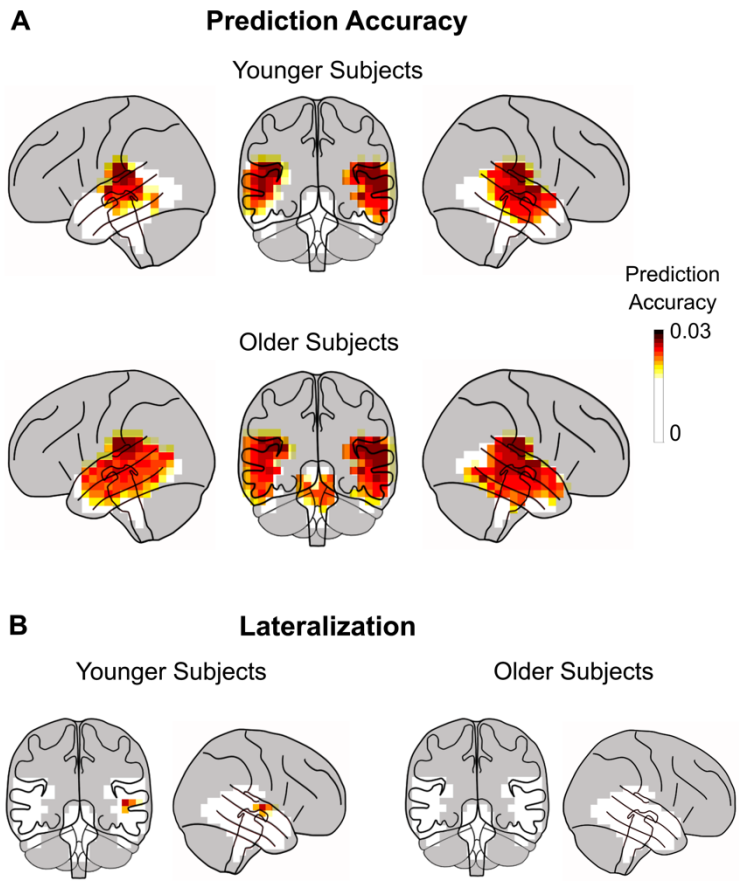

Figure 3.2. Prediction accuracy of volume source localized TRFs. A. Prediction accuracy using the TRF model for each voxel in the volume source space ROIs (non-gray regions) averaged across subjects. Only ROI voxels for which model prediction accuracy significantly increased over the noise model are plotted ($p < 0.05$, corrected). The prediction accuracy is larger in cortical areas than in subcortical areas. Plots are of the maximum intensity projection, with an overlay of the brain. When taking into account expected MEG volume source localization leakage, these results are consistent with the response originating solely from cortical areas and with a right hemispheric bias. **B.** An area in the right hemisphere near the auditory cortex is significantly more predictive than the left hemisphere, but only in the younger subjects.

The TRFs at each source voxel are represented by a 3-dimensional current vector that varies over the time lags. Hence for each voxel and time lag, the amplitude of the TRF vector for the true model was tested for significance against the average of the noise models across subjects using a one-tailed test with paired sample *t*-values and TFCE. The TRFs for the envelope modulation predictor in the cortical ROI were significant (younger $t_{max} = 5.38$, $p < 0.001$; older $t_{max} = 4.69$, $p < 0.001$) starting at a time lag of 23 ms, and ending at 63 ms, with an average peak latency of 40 ms (see Fig. 3.3A). The TRF current dipoles oscillate with alternating direction between successive amplitude peaks. However, in all subsequent TRF plots, the TRF amplitude is shown, and not signed current values, and hence signal troughs and peaks both appear as peaks. The subcortical ROI was also analyzed in a similar manner and the TRF showed significance in a much smaller time range of 31–35 ms only for older subjects (younger $t_{max} = 2.96$, $p > 0.13$; older $t_{max} = 3.69$, $p < 0.01$) (see Fig. 3.3B). There was no significant difference in amplitudes between younger and older subjects (cortical ROI $t_{max} = 3.7$, $t_{min} = –3.38$, $p > 0.18$; subcortical ROI $t_{max} = 3.05$, $t_{min} = –3.39$, $p > 0.45$). The TRF responses oscillate at a frequency of ~80 Hz (see below for a more detailed spectral analysis). The amplitude of these TRFs was significantly larger in voxels in Heschl's gyrus than in the subcortical ROI (two-tailed test with paired sample t-values on the $l2$ norm of the TRFs across subjects: younger $t = 3.51$, $p = 0.003$; older $t = 4.52$, $p < 0.001$). Since the subcortical TRFs also have a similar latency and shape to the cortical TRFs, and because a latency of 23 to 63 ms is late for a subcortical response, these subcortical TRFs are consistent with artifactual leakage from the cortical TRFs

due to the spatial spread of MNE source localization. Simulated volume source estimates for current dipoles originating only in Heschl's gyrus generated a spatial distribution of TRF directions consistent with the experimental data (see Appendix), i.e. the spatial spread of MNE localized cortical responses resulted in apparent TRF vectors even in the subcortical ROI. These results indicate that the response originates predominantly from cortical regions.

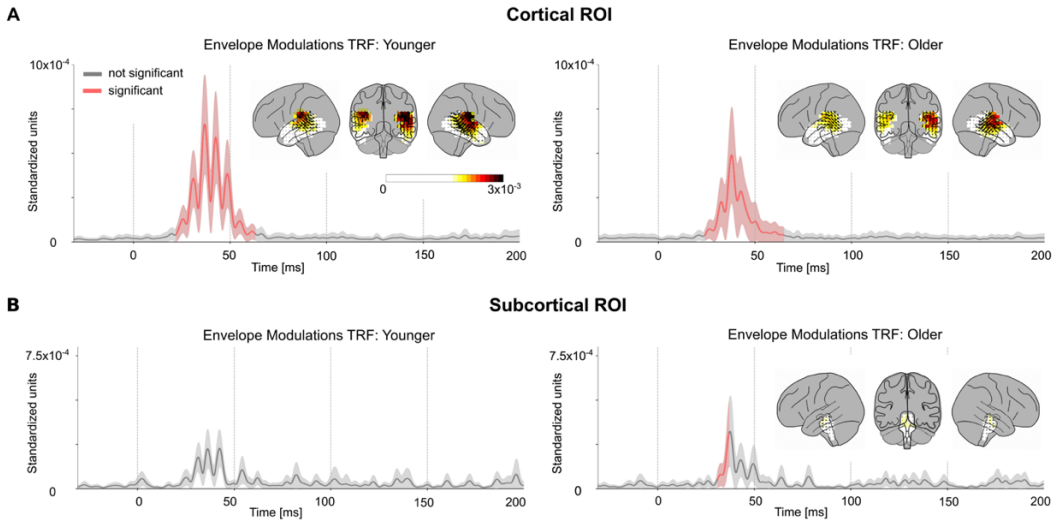

Figure 3.3. Volume source localized envelope modulation TRFs. The amplitude of the TRF vectors for the envelope modulation predictor averaged across voxels in the ROI, and, the mean ± (standard error) across subjects is plotted in the cortical (A) and subcortical (B) ROIs. Red curves are time points when the TRF showed a significant increase in amplitude over noise. The TRF was resampled to 2000 Hz for visualization purposes. The TRF shows a clear response with a peak latency of ~40 ms. The distribution of TRF vectors in the brain at each voxel at the time with the maximum response are plotted as an inset for each TRF, with color representing response strength and the arrows representing the TRF directions. The color bar represents the response strength for

all 4 brain insets. The response oscillates around a frequency of ~80 Hz and is much stronger in the cortical ROI compared to the subcortical ROI. Note that since only the TRF amplitude is shown, and not signed current values, signal troughs and peaks both appear as peaks. In the original, signed TRFs, the current direction alternates between successive amplitude peaks. The latency and amplitude of the response suggests a predominantly cortical origin.

3.4.2. Responses to the envelope modulation and the carrier

Next, the neural response to the carrier was compared with that to the envelope modulation. The carrier TRF was also tested for significance using a corresponding noise model (as employed above). The carrier TRF showed weak responses that were only significant in the cortical ROI between 33–51 ms (younger $t_{max} = 3.70$, $p = 0.042$; older $t_{max} = 4.7$, $p < 0.001$) (see Fig. 3.4A, B). Although the carrier and envelope modulation predictors are correlated ($r = –0.42$), the TRF analysis is able to separate the contributions of these two predictors remarkably well. Two-tailed paired sample *t*-values and TFCE were used to test for a significant increase of the *l*2 norm of the envelope modulation TRF when compared to the carrier TRF in a time window of 20–70 ms in the cortical ROI (see Fig. 3.5A). This test was significant for both younger ($t_{max} = 4.38$, $p = 0.002$) and older ($t_{max} = 3.63$, $p = 0.017$) subjects. However, this test did not find a significant increase in the envelope modulation TRF over the carrier TRF in the subcortical ROI for either younger ($t_{max} =0.045$, $p > 0.32$) or older subjects ($t_{max} = 0.89$, $p > 0.36$). Since the TRF analysis allows both stimulus predictors to directly

compete for explaining response variance, the results strongly indicate that the response is primarily due to the envelope modulation over the carrier.

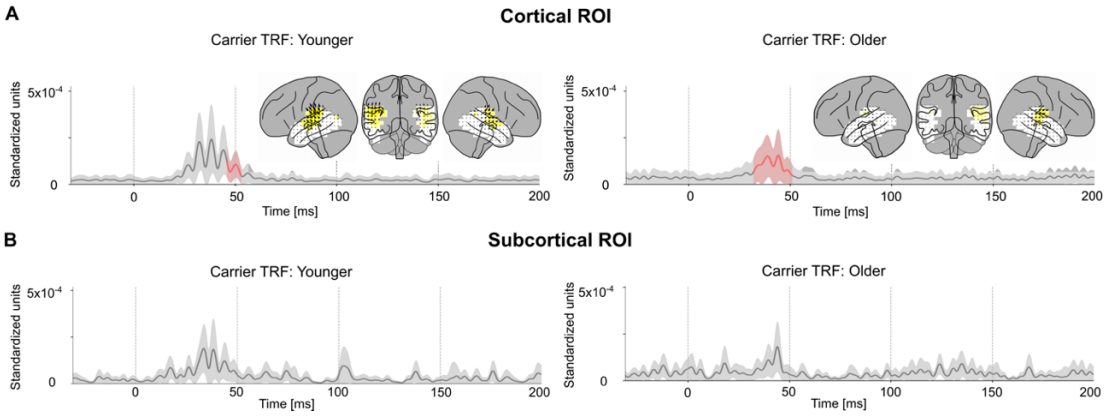

Figure 3.4. Volume source localized carrier TRFs. The amplitude of the TRF vectors for the carrier predictor averaged across sources. Mean ± (standard error) across subjects is shown, analogous to Fig. 3.3. For comparison, the axis and color scale are identical to that in Fig. 3.3. The TRF shows a weaker response compared to the case of envelope modulation, with a peak latency of ~40 ms, that is significant in the cortical ROI for both groups, and over a longer time interval for older subjects. Comparison with Fig. 3.3 suggests that the high gamma response is dominated by the envelope modulation over the carrier.

3.4.3. Age-related differences

Statistical tests were performed for age-related differences between older and younger subjects on both the prediction accuracy and the TRFs. Two-tailed tests of prediction accuracy with independent sample *t*-values and TFCE indicated no significant difference (cortical ROI $t_{max} = 1.17$, $t_{min} = -2.72$, $p > 0.44$; subcortical ROI $t_{max} = -0.78$, $t_{min} = -1.37$, $p > 0.38$). Similarly, no voxels or time points were

significantly different in either the envelope modulation TRF (cortical ROI t_{max} = 3.7, t_{min} = –3.38, p > 0.18; subcortical ROI t_{max} = 3.05, t_{min} = –3.39, p > 0.45) or the carrier TRF (cortical ROI t_{max} = 3.34, t_{min} = –3.89, p > 0.25; subcortical ROI t_{max} = 2.69, t_{min} = –3.10, p > 0.18). In addition, the cortical ROI TRFs showed no significant differences across age groups in peak latency (envelope modulation TRF t_{max} = 1.82, t_{min} = –2.62, p > 0.5; carrier TRF t_{max} = 2.79, t_{min} = –2.32, p > 0.53). An additional analysis was performed using surface source space TRFs as described in detail in the Appendix. Both high (70–200 Hz) and low (1–10 Hz) frequency TRFs were computed in surface source space, and model prediction accuracy was assessed with an ANOVA with factors TRF frequency and age. The ANOVA showed a significant frequency × age interaction ($F_{1,38}$ = 6.46, p = 0.015), suggesting that age related differences are indeed not consistent across high and low frequency responses (detailed results in Appendix), i.e. present at low but not at high frequencies.

3.4.4. Pitch analysis

To further understand the contributions of these predictors to the TRF oscillations, the frequency spectrum of the TRFs and the predictors were compared (see Fig. 3.5B). The frequency spectrum of the average TRFs showed a broad peak centered near 80 Hz for both predictors and both age groups (envelope TRF spectral peak mean = 81 Hz, std = 5 Hz; carrier TRF spectral peak mean = 82 Hz, std = 8 Hz). In contrast, the spectral peak of the predictor variables was near 110–120 Hz for the carrier, and near 70–75 Hz for the envelope modulation. Since the TRF peak frequency did not match

the peak power in either of the predictors, a further analysis was performed after separating the stimulus into high- and low-pitch time segments (see Methods). This resulted in a model with 4 predictors and their corresponding TRFs: high/low-pitch envelope modulation and high/low-pitch carrier. The low-pitch envelope modulation TRFs and low-pitch carrier TRFs are broadly similar to those of the earlier analysis (see Fig. 3.6). These TRFs show more significant regions than the previous analysis, although the two models (one with 2 predictors, the other with 4 predictors) cannot be directly compared since an increased number of predictors has more degrees of freedom and allows for the model to predict more of the signal. The TRF amplitudes were significantly larger in the low pitch TRFs when compared to the high pitch TRFs (see Fig. 3.5C; envelope modulation $t_{max} = 7.6$, $p < 0.001$; carrier $t_{max} = 3.78$, $p = 0.013$). In addition, the spectra of the low pitch TRFs peak near 80 Hz similar to the low pitch predictors (envelope TRF spectral peak mean = 81 Hz, std = 6 Hz; carrier TRF spectral peak mean = 82 Hz, std = 4 Hz), while the high pitch TRFs do not have a clear peak (see Fig. 3.5D). This suggests that the TRF oscillation is driven mainly by the segments of the stimulus with pitch below 100 Hz, and that responses to stimulus pitches above 100 Hz are not easily detected by this analysis.

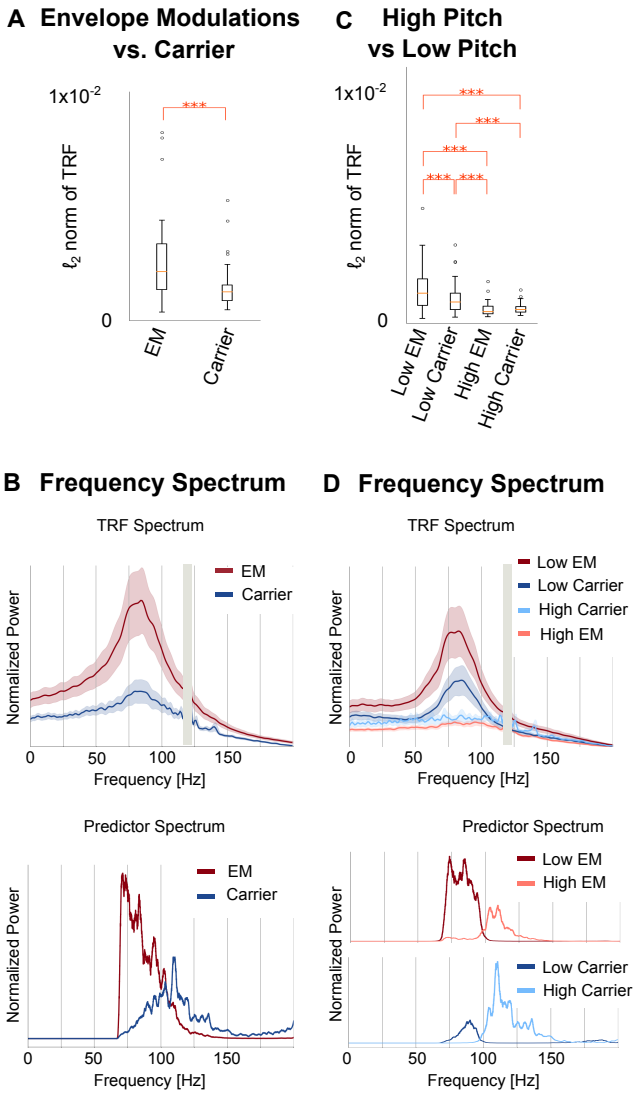

Figure 3.5. Comparison of responses to the envelope modulation and to the carrier. A. The ℓ_2 norm of the TRF between 20 ms and 70 ms was larger in the envelope modulation TRF than the carrier TRF (*** $p < 0.001$). Boxplots after combining both age groups are shown. **B**. The frequency spectrum of the TRF reveals that the oscillation has a broad peak around 80 Hz (vertical gray bars denote a narrow frequency band excluded from analysis because of 120 Hz line noise). In contrast, the predictors' peaks are displaced in frequency from the TRF peak, either well below (for the envelope modulation) or well above (for the carrier). Note that the sharp cutoff in the envelope

modulation spectrum at 70 Hz arises from the bandpass filter used in analysis: without the bandpass filter the spectrum would continue rising toward lower frequencies. **C.** The ℓ_2 norm of the TRF for the pitch-separated model between 20 ms and 70 ms was larger in the low pitch TRFs than the high pitch TRFs for both envelope modulation and carrier (*** $p < 0.001$). **D**. The frequency spectrum of the low-pitch TRF has a peak around 80 Hz, while the high pitch TRF does not show any peaks. This suggests that the TRF is dominantly driven by the low-pitch segments of the speech waveform. The spectra of the corresponding high and low pitch predictors are also shown, highlighting the clear separation of the spectra at the median pitch frequency of 98 Hz.

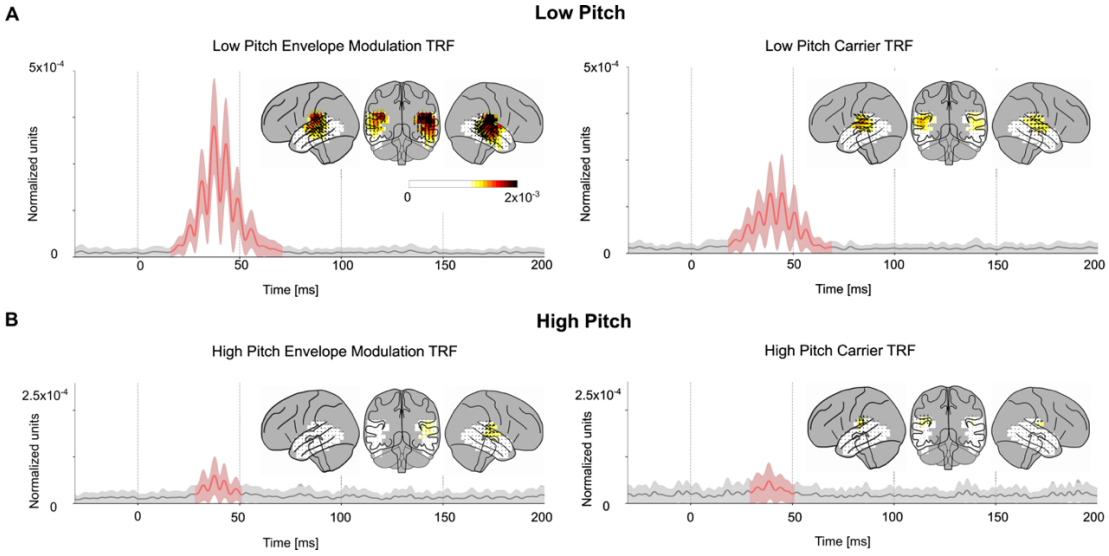

Figure 3.6. Pitch-separated TRFs. The amplitude of the low pitch TRF (A) and high pitch TRF (B) vectors for both the envelope modulation and carrier predictors averaged across sources. Mean ± (standard error) across subjects is shown, analogous to Fig. 3.3, 3.4. The axis and color scale are smaller than those in Fig. 3.3, 3.4, since the pitch separated TRFs are each based on a subset of the stimulus predictors, and hence have weaker amplitudes. All four TRFs show significant regions around 40 ms, but the low pitch envelope modulation TRF is the strongest, followed by the low pitch carrier TRF. This indicates that the high gamma response is time-locked to the low pitch segments of the speech stimulus.

3.5. Discussion

In this study, we investigated high gamma time-locked responses to continuous speech measured using MEG. Such responses were found, and their volume source localized TRFs provided evidence that these responses originated from cortical areas with a peak response latency of approximately 40 ms. The responses showed a significant right hemispheric asymmetry. These responses oscillate with a frequency of approximately 80 Hz and track the low pitch segments of the speech stimulus. We also showed that the response is significantly stronger to the envelope modulation than the carrier. Surprisingly, there were no significant age-related differences in response amplitude, latency, localization or predictive power. This is in contrast to age-related differences seen in both the subcortical EEG FFR (younger > older) and the cortical low frequency TRF (older > younger).

3.5.1. MEG sensitivity to high gamma responses

MEG signals are known to have poor SNR at high frequencies ($\gtrsim$100 Hz) (Hansen et al., 2010). The MEG signal is an average over a large population of neurons, and hence detection of population level high gamma responses requires precise (within a few ms) phase synchrony across these populations (Hämäläinen et al., 1993). However, the cortical sources which MEG would otherwise be sensitive to rarely phase synchronize across a large population at these high gamma ranges, leading to poor

neural SNR for these high gamma ranges (reflected in our results by the small correlation values between actual responses and model predictions). The implications of this for our study are twofold. Firstly, conclusions regarding the intrinsic properties of high gamma responses to speech are limited by these methodological constraints on the MEG signal. Our results only show that there are significant cortical responses at ~80 Hz, but do not rule out higher frequency cortical responses, or subcortical responses, that may be buried in poor SNR. Conversely, however, it is somewhat surprising that using such a simple experimental paradigm, with short duration continuous natural speech, it is possible to reliably detect such MEG responses using a TRF model.

3.5.2. MEG sensitivity to deep sources

Gradiometer-based MEG is physically constrained to be less sensitive to deep structures, typically resulting in such subcortical MEG responses being up to 100 times weaker than cortical responses at equivalent current strengths (Attal et al., 2007; Hillebrand and Barnes, 2002). Several source localization techniques have been proposed to correct for this inherent bias towards cortical sources (Dale et al., 2000; Krishnaswamy et al., 2017; Pascual-Marqui, 2002). Some studies were able to resolve MEG responses to the hippocampus (Cornwell et al., 2012), amygdala (Balderston et al., 2014; Cornwell et al., 2008; Dumas et al., 2013) and thalamus (Roux et al., 2013). Prior work has also been done using MEG for measuring brainstem responses (Coffey et al., 2016; Parkkonen et al., 2009). These studies show that MEG can be used to

localize sources in subcortical areas given a large number of repetitions or specialized experimental paradigms. However, some of these studies used magnetometer-based MEG, which is more sensitive to deep sources than gradiometer based MEG (Lopes da Silva and van Rotterdam, 2005). In addition, resolving several sources with MEG is more complicated than localizing an isolated source due to the non-unique nature of distributed inverse solutions (Lütkenhöner, 2003). In our study, we used such a distributed source localization and a short experimental paradigm (without many stimulus repetitions) and found responses dominated by cortical sources. Simulation results suggested that the small amount of activation associated with the brainstem is more easily explained as an artifact of source localization leakage from cortical sources. On the other hand, although these results do not identify responses from subcortical regions, this does not imply at all that such responses are absent in the auditory system. Brainstem responses to continuous speech have been detected using EEG (Maddox and Lee, 2018), and it is entirely possible that high gamma subcortical responses to speech may also be detected in MEG by other experiments with higher SNR, different analysis methods or MEG systems that are more sensitive to deep sources.

3.5.3. Cortical FFRs and high gamma TRFs

The high gamma TRF is not directly analogous to the FFR because, among other reasons, it is not an average over several repetitions of simple stimuli, but is instead a weighted average over longer time. However, the TRFs measured here indeed show a measure of high gamma time-locking that can be compared to the FFR. Cortical FFRs

to repeated single speech syllables have been measured in MEG (Coffey et al., 2016) and EEG (Bidelman, 2018; Coffey et al., 2017b). Our work shows that cortical TRFs contain significant responses up to 62 ms, comparable to the long-lasting explanatory power of the auditory cortex ROI in Coffey et al., 2016. These TRFs are also predominantly from auditory cortex, centered around Heschl's gyrus, and right lateralized similar to the MEG FFR (Coffey et al., 2016). However, some studies have demonstrated that the contribution of cortical sources to the FFR as measured with EEG is weaker than when measured with MEG (Ross et al., 2020), and rapidly decreases for harmonics above 100 Hz (Bidelman, 2018). In fact, while subcortical FFR is measurable with EEG for harmonics up to 1000 Hz, there were no cortical contributions to the FFR above 150 Hz (Bidelman, 2018). Unsurprisingly our results confirm that the cortical sources dominate the MEG response at frequencies near 100 Hz.

3.5.4. Comparison of responses to the envelope modulation vs. the carrier

The subcortical FFR is typically analyzed by averaging across stimulus presentations of opposite polarity, which results in responses driven mainly by the stimulus envelope and other even-order nonlinearities (Lerud et al., 2014). However, studies have also analyzed the FFR by subtracting the responses to stimulus presentations of opposite polarity (Aiken and Picton, 2008), which is driven mainly by the carrier and odd-order nonlinearities. Hence both the envelope and the carrier modulate responses across the auditory pathway. Unlike FFR analysis, the TRF analysis used in this study is well suited to disentangle the contributions of different

features of the stimulus to the neural response, since it allows each stimulus representation to directly compete to explain the response variance. Our results found significant time-locked high gamma cortical responses to continuous speech for both envelope modulation and carrier, but these responses were predominantly driven by the envelope modulation over the carrier. This could be related to the perceptual phenomenon that modulation of the speech spectrum above 300 Hz is more behaviorally relevant for speech understanding, and more resistant to background noise, than the carrier below 200 Hz (Assmann and Summerfield, 2004). Slow evoked responses in auditory cortex are also sensitive to fine-structure acoustic features such as pitch and timbre (Roberts et al., 2000), and the auditory cortical response to the slowly varying envelope of speech is likewise modulated by the spectrotemporal fine structure of the stimulus (Ding et al., 2014).

3.5.5. High gamma TRF is driven by low pitch segments of the speech

The TRF response oscillates with a peak frequency of approximately 80 Hz, and is well time-locked to the segments of speech where the pitch is below 100 Hz. Cortical auditory phase locked responses to simple sounds have been measured using MEG (Coffey et al., 2016; Hertrich et al., 2004; Schoonhoven et al., 2003) at frequencies of up to 111 Hz. For continuous speech stimuli, such phase locked responses could reflect a cortical mechanism that represents complex speech features such as modulations in vowel formants, using fluctuations in the fundamental frequency domain of natural speech. Our pitch analysis showed that the response strongly locks to pitch frequencies

below 100 Hz, but not above 100 Hz. This agrees with other studies that show a bias in cortical phase-locking towards lower frequencies (Bidelman, 2018; Ross et al., 2000; Schoonhoven et al., 2003).

3.5.6. Right lateralization of responses

The TRF model prediction accuracy was significantly right lateralized in younger subjects. The lack of significant right lateralization among older subjects may not indicate an age-related lateralization difference, but rather a lack of statistical power, since the lateralization was not significantly different across age groups. However, similar lateralization differences across age groups have been found for 80 Hz ASSR (Goossens et al., 2016). Stronger responses in the right auditory cortex have been observed for ASSR using EEG (Ross et al., 2005) and MEG (Hertrich et al., 2004) as well as in cortical FFRs using MEG (Coffey et al., 2016). This agrees with prior studies showing that right auditory cortex is specialized for early tonal processing and pitch resolution (Cha et al., 2016; Hyde et al., 2008; Zatorre, 1988). Both this right hemispheric bias, and the relatively short peak latency of 40 ms of our TRFs suggest that these cortical high gamma responses are due to early auditory processing of acoustic periodicity. However, some studies have also suggested that increased cortical folding in left auditory cortex could lead to a cancellation of MEG signals in the left hemisphere, which could lead to a similar right-ward bias in the absence of functional lateralization (Shaw et al., 2013).

3.5.7. Absence of age-related differences

Temporal precision and synchronized activity decreases in the auditory system with age and is characterized by age related differences in both subcortical and cortical responses. Older adults have subcortical FFR responses with smaller amplitudes, longer latencies and reduced phase coherence, which could be due to an excitation-inhibition imbalance or a lack of neural synchrony (Hornickel et al., 2012). In a surprising reversal, MEG and EEG studies have revealed that older adults have larger slow (below ~10 Hz) cortical responses than younger adults (Alain et al., 2014; Bidelman et al., 2014; Herrmann et al., 2016), that result in better prediction accuracy for reverse correlation methods (Decruy et al., 2019; Presacco et al., 2016a, 2016b). Animal studies suggest that this opposite effect could be due to cortical compensatory central mechanisms (Chambers et al., 2016; Salvi et al., 2017) or lack of inhibition (Caspary et al., 2008; Villers-Sidani et al., 2010). Another possibility is the recruitment of additional neural areas for redundant processing (Brodbeck et al., 2018b; Peelle et al., 2010). Contrary to both these cases, we found no significant age-related differences in high gamma cortical responses, although this might be due to a lack of statistical power (see Ross et al., 2020). An ANOVA with factors TRF frequency and age suggested that the difference in low frequency responses among older and younger adults is not preserved for high gamma responses (see Appendix). These results suggest that high gamma cortical responses do not show a clear difference with age. The high gamma MEG TRF reflects fine-grained time-locked neural activity, like a subcortical FFR, but arising from cortical areas. It is possible that older adults' exaggerated

responses in cortical areas and the lack of neural synchrony at high frequencies (as seen in subcortical FFRs) affect their high gamma MEG responses in opposite directions and obscure what would otherwise be detectable age-related differences.

3.5.8. Neural mechanisms for the MEG high gamma response

Given that MEG records the aggregate response over a large population of neurons, the specific origins of high gamma time-locked responses are not readily apparent. It is possible that the high gamma TRF reflects the effects of several processing stages along the auditory pathway, similar to the FFR. Electrocorticography (ECoG) studies have seen cortical phase-locked activity at these high rates (Nourski et al., 2014; Steinschneider et al., 2013). However, cortical phase-locking at the individual neuron level drastically reduces with increasing frequency (Lu et al., 2001), and hence cortical neurons may not be the sole contributor to these high gamma responses.

Such phase locked auditory activity is compatible with the spiking output of the Medial Geniculate Body (MGB) (Miller et al., 2002), which provides input to early auditory cortical areas. The MEG signal is dominantly driven by dendritic currents (that give rise to the Local Field Potential) (Hämäläinen et al., 1993), and hence these high gamma responses may be due to the inputs from the MGB into auditory cortex. Prior work has shown that auditory cortex is able to transiently time-lock to continuous acoustic features with surprisingly high temporal precision of the order of milliseconds (Elhilali et al., 2004). Time-locked inputs from MGB may provide a neural substrate for such precise transient temporal locking to stimulus features. Direct correspondences

with age-related changes in thalamus from animal work are limited (Caspary and Llano, 2019), and hence it is unclear if time-locked high gamma spiking activity in MGB animal models would be similar across age. However, invasive neural recordings could help to disentangle the opposite effects of aging in the brainstem and the cortex seen with MEG and EEG, leading to a better understanding of time-locked responses in the aging auditory pathway.

3.6. Conclusion

In this study, we found high gamma time-locked responses to continuous speech, using MEG, that localized to auditory cortex, occurred with a peak latency of approximately 40 ms, and were stronger in the right hemisphere. We showed that TRF analysis could be used to reliably separate the contributions of several stimulus features to this response. The response function showed oscillations at approximately 80 Hz, predominantly driven by the envelope modulations during the segments of the speech where the pitch is below 100 Hz. Such high gamma time-locked responses may originate from the thalamic inputs to cortical neurons. These responses can be reliably detected in MEG using natural speech stimuli even of short duration, allowing TRF analysis to be employed to investigate auditory processing of speech from acoustics to semantics under several stimulus conditions in the same experiment. Furthermore, there were no significant age-related differences in these high gamma responses, unlike in both the low frequency cortical TRFs or the subcortical FFRs. Hence both the neural

origin and the frequency domain must be considered when investigating age-related changes in the auditory system.

3.7. Appendix

3.7.1. Simulation of spatial spread of distributed source localization

Distributed neural source localization methods for MEG, such as MNE, result in a substantial amount of spatial spread. In order to characterize this spread, a dipole was simulated on Heschl's gyrus perpendicular to the pial surface of the 'fsaverage' brain using the 'ico-4' surface source space. The dipole was then projected to sensor space, and MNE source localization with dSPM was performed to project it back onto the volume source space (see Fig. 3.A1). The peak of the activity shows a broad spread around Heschl's gyrus but also some small activity in other parts of temporal lobe and even in the brainstem. This supports the claim that high gamma responses seen at the brainstem in our study are attributable to leakage from cortical areas.

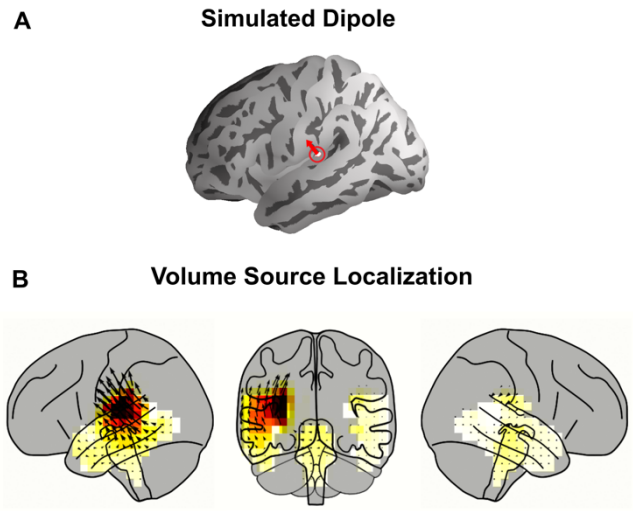

Figure 3.A1. Simulation of spatial spread of volume source localization. A. One dipole in Heschl's gyrus was simulated. **B.** Volume source localization of the dipole after it was projected to sensor space. The cortical and subcortical ROIs are shown and artifactual leakage is seen in the brainstem voxels.

3.7.2. Surface source space TRF methods and results

Cortical surface source space estimation was performed using the 'ico-4' source space, which consists of a fourfold icosahedral subdivision of the white matter surface of cortex with dipoles oriented normal to the surface. The 'aparc' parcellation was used to select dipoles in the temporal lobe for further analysis. In this surface source space analysis, current dipoles have a fixed orientation normal to the surface, and hence the TRF consists only of signed scalar amplitude variations with time. The Pearson correlation between the actual and predicted neural response was used as a measure of prediction accuracy for each neural source. For statistical tests, the TRFs and the correlation values were first rectified and then spatially smoothed using a Gaussian

window with a standard deviation of 5 mm. The rectified, smoothed TRF of the true model was compared to the average of that of the three noise models using the same one tailed test with paired sample *t*-values and the TFCE procedure outlined in Methods.

Lateralization tests were performed to check for hemispheric asymmetry. The correlation values at each neural source in both left and right hemisphere were morphed onto the right hemisphere of the 'fsaverage_sym' brain as described in Brodbeck et al. (2018a). This brain model is symmetric in left and right hemispheres, allowing for comparisons between corresponding neural sources in both hemispheres. As before, these correlation coefficients were spatially smoothed using the same Gaussian window. After morphing, the correlation values of the average noise model were subtracted from that of the true model and a two-tailed test with paired sample *t*-values and TFCE was used to assess for significant differences in each of the corresponding left and right current dipoles.

TRFs were estimated using the cortical surface source space for neural sources in the temporal lobe, using both the envelope modulation and the carrier predictors in a competing model. Both predictors were time-shifted to generate noise models. All surface space results were similar to volume source space results. The prediction accuracies and TRFs are shown in Fig. 3.A2, Fig. 3.A3. The prediction accuracies were right lateralized but only in younger subjects ($t_{max} = 4.6$, $p = 0.008$). The TRFs showed a significant response in the range of 19–67 ms for the envelope modulation and 23–57 ms for the carrier. The envelope modulation TRF was stronger than the carrier TRF

using the same tests as in the volume source space (younger t_{max} = 5.27, $p < 0.001$; older t_{max} = 3.46, $p = 0.03$). There were no age-related differences in surface source space analyses (prediction accuracy t_{max} = 2.41, t_{min} = –2.99, $p > 0.29$; maximum amplitude of envelope modulation TRF t_{max} = 2.40, t_{min} = –2.47, $p > 0.68$; maximum amplitude of carrier TRF t_{max} = 1.79, t_{min} = –3.07, $p > 0.54$).

In addition, low frequency TRFs were also estimated to compare age-related differences in both frequency domains. The stimulus representation for this model was the Hilbert envelope of the speech waveform filtered at 1–10 Hz with a logarithmic nonlinearity applied. The MEG data was also filtered at 1–10 Hz and TRFs were estimated using the surface source space. The resulting TRFs were as expected from prior work (Brodbeck et al., 2018b), with older subjects showing significantly higher reconstruction accuracies (t_{max} = 0.93, t_{min} = –3.45, $p = 0.022$). The increase in model prediction accuracies above the noise, for the high frequency TRF and the low frequency TRF were averaged across neural sources per subject, and a TRF frequency by age ANOVA was performed. Results indicated a significant interaction of TRF frequency × age ($F_{1,38}$ = 6.46, p = 0.015) and significant main effects of TRF frequency ($F_{1,38}$ = 216.58, $p < 0.001$) and age ($F_{1,38}$ = 4.83, $p = 0.034$). This suggests that age-related changes are not consistent across low and high frequency responses, in further agreement with all the above results.

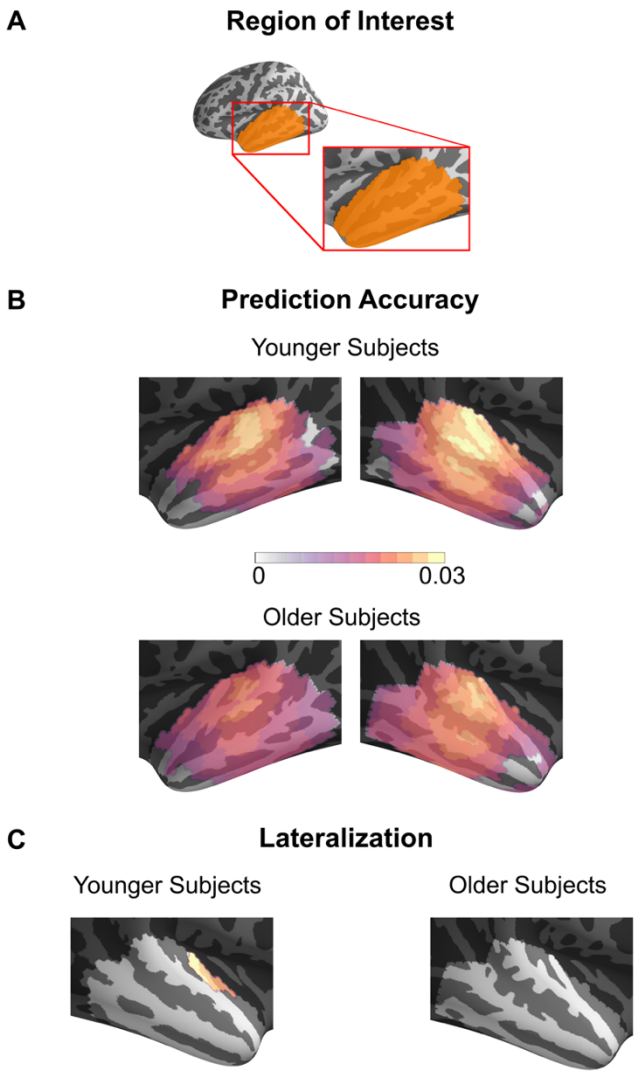

Figure 3.A2. Prediction accuracy of surface source space TRFs. Pearson correlation coefficients between the actual and predicted response using the TRF model for each source in the surface source space ROI averaged across subjects are shown on an inflated brain. Only the voxels showing a significant increase in prediction accuracy over the noise model are plotted. Although most neural sources are significantly predictive, the prediction accuracy is larger in areas near core auditory cortex. A region in auditory cortex is significantly more predictive in the right hemisphere than the left, but only in younger subjects.

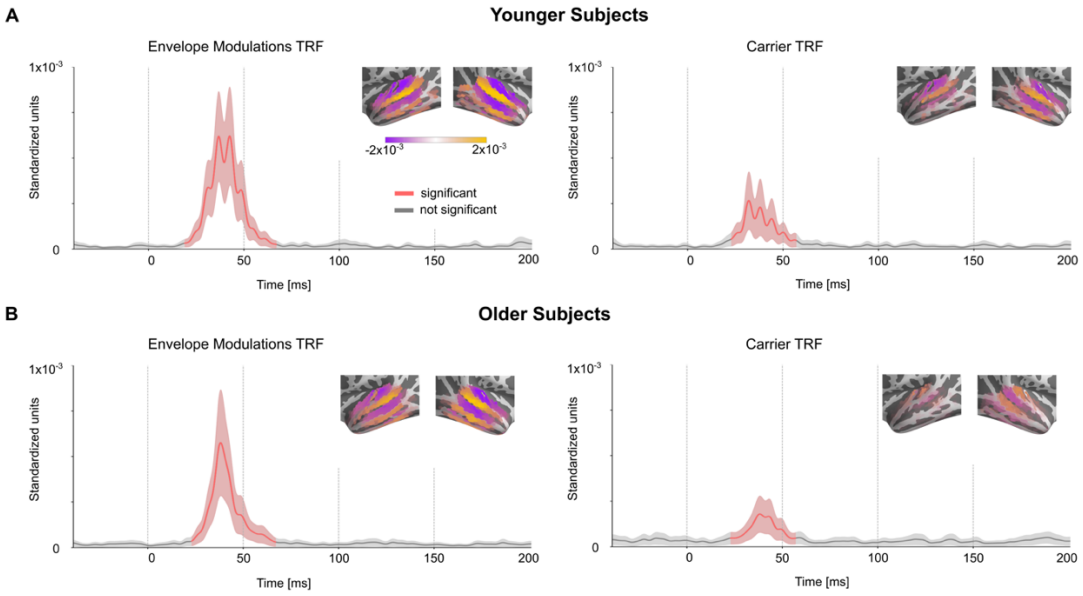

Figure 3.A3. Surface source space TRFs. The amplitude of the TRFs for the competing model for both predictors averaged across neural sources and masked by significance against the noise model. Mean ± (standard error) across subjects is shown. The distribution of current dipoles in the temporal lobe ROI at the peak of the response is shown as an inset. Unlike the volume source space, the surface source space comprises of current dipoles with fixed orientation normal to the cortical surface. The signed magnitudes of these fixed direction dipoles are plotted on the surface, allowing for positive (orange) and negative (purple) values for outward and inward directions.

Chapter 4

Cortical Processing of Arithmetic and Simple Sentences in an Auditory Attention Task

This work has been published as

Kulasingham, J.P., Joshi, N.H., Rezaeizadeh, M.*, Simon, J.Z., 2021. Cortical Processing of Arithmetic and Simple Sentences in an Auditory Attention Task.*

J. Neurosci. 41, 8023–8039.

https://doi.org/10.1523/JNEUROSCI.0269-21.2021

**contributed equally to this work*

4.1. Abstract

Cortical processing of arithmetic and of language rely on both shared and task-specific neural mechanisms, which should also be dissociable from the particular sensory modality used to probe them. Here, spoken arithmetical and non-mathematical statements were employed to investigate neural processing of arithmetic, compared to general language processing, in an attention-modulated cocktail party paradigm. Magnetoencephalography (MEG) data were recorded from 22 human subjects listening to audio mixtures of spoken sentences and arithmetic equations while selectively attending to one of the two speech streams. Short sentences and simple equations were presented diotically at fixed and distinct word/symbol and sentence/equation rates. Critically, this allowed neural responses to acoustics, words, and symbols to be

dissociated from responses to sentences and equations. Indeed, the simultaneous neural processing of the acoustics of words and symbols was observed in auditory cortex for both streams. Neural responses to sentences and equations, however, were predominantly to the attended stream, originating primarily from left temporal, and parietal areas, respectively. Additionally, these neural responses were correlated with behavioral performance in a deviant detection task. Source-localized Temporal Response Functions revealed distinct cortical dynamics of responses to sentences in left temporal areas and equations in bilateral temporal, parietal, and motor areas. Finally, the target of attention could be decoded from MEG responses, especially in left superior parietal areas. In short, the neural responses to arithmetic and language are especially well segregated during the cocktail party paradigm, and the correlation with behavior suggests that they may be linked to successful comprehension or calculation.

Significance Statement

Neural processing of arithmetic relies on dedicated, modality independent cortical networks that are distinct from those underlying language processing. Using a simultaneous cocktail party listening paradigm, we found that these separate networks segregate naturally when listeners selectively attend to one type over the other. Neural responses in the left temporal lobe were observed for both spoken sentences and equations, but the latter additionally showed bilateral parietal activity consistent with arithmetic processing. Critically, these responses were modulated by selective attention and correlated with task behavior, consistent with reflecting high-level processing for

speech comprehension or correct calculations. The response dynamics show task-related differences that were used to reliably decode the attentional target of sentences or equations.

4.2. Introduction

Comprehension and manipulation of numbers and words are key aspects of human cognition and share many common features. Numerical operations may rely on language for precise calculations (Pica et al., 2004; Spelke and Tsivkin, 2001) or share logical and syntactic rules with language (Houdé and Tzourio-Mazoyer, 2003). During numerical tasks, frontal, parietal, occipital and temporal areas are activated (Arsalidou and Taylor, 2011; Dastjerdi et al., 2013; Dehaene et al., 2004, 2003; Harvey et al., 2013; Harvey and Dumoulin, 2017; Maruyama et al., 2012; Menon et al., 2000). Bilateral intraparietal sulcus (IPS) is activated by presenting numbers using Arabic or alphabetical notation (Pinel et al., 2001) or speech (Eger et al., 2003). Posterior parietal and prefrontal areas are activated for both arithmetic and language (Bemis and Pylkkänen, 2013; Göbel et al., 2001; Price, 2000; Venkatraman et al., 2006; Zarnhofer et al., 2012). However, some cortical networks activated by numerical stimuli (e.g., IPS), differ from those underlying language processing, even when the stimuli are presented using words (Amalric and Dehaene, 2016, 2019; Monti et al., 2012; Park et al., 2011). Lesion studies (Baldo and Dronkers, 2007; Dehaene and Cohen, 1997; Varley et al., 2005) further provide evidence that the neural basis of numerical

processing is distinct from that of language processing (Amalric and Dehaene, 2018; Gelman and Butterworth, 2005).

The dynamics of these neural processes have also been investigated. Evoked responses to arithmetic have been found in parietal, occipital, temporal and frontal regions (Iguchi and Hashimoto, 2000; Iijima and Nishitani, 2017; Jasinski and Coch, 2012; Kou and Iwaki, 2007; Ku et al., 2010; Maruyama et al., 2012). Arithmetic operations can even be decoded from such responses (Pinheiro-Chagas et al., 2019). Speech studies differentiate early auditory evoked components from later components reflecting linguistic and semantic processing in temporal, parietal and frontal regions (Baggio and Hagoort, 2011; Koelsch et al., 2004; Lau et al., 2008; Obleser et al., 2003, 2004). Linear models of time-locked responses to continuous speech called Temporal Response Functions (TRFs) have also revealed dynamical processing of linguistic features.

To investigate cortical processing of spoken language and arithmetic, we utilize a technique pioneered by Ding et al. (2016) of presenting isochronous (fixed rate) words and sentences. There, the single syllable word rate, also the dominant acoustic rate, is tracked strongly by auditory neural responses, as expected. However, cortical responses also strongly track the sentence rate, completely absent in the acoustics, possibly reflecting hierarchical language processing (Jin et al., 2020; Luo and Ding, 2020; Sheng et al., 2018). When subjects selectively attend to one speech stream among several, in a ‘cocktail party paradigm’, the sentence rate is tracked only for the attended speaker (Ding et al., 2018). Similarly, cocktail party studies using TRFs show early

auditory responses irrespective of attention, and later attention-modulated responses to higher order speech features (Brodbeck et al., 2018c; Ding and Simon, 2012b). Attention modulates activation related to numerical processing as well (Castaldi et al., 2019).

Here, magnetoencephalography (MEG) is used to study the cortical processing of short spoken sentences and simple arithmetic equations, presented simultaneously at fixed sentence, equation, word and symbol rates, in an isochronous cocktail party paradigm. This study is motivated by several questions, of increasing complexity. The most basic is whether isochronously presented equations allow segregation of equation-level from symbol-level neural processing in the frequency domain. We demonstrate strong evidence for this segregation. The next level is whether equation- and sentence-level processing show shared or distinct cortical activity areas. We demonstrate evidence for both: shared activity in the left temporal lobe, and distinct equation processing in bilateral IPS and occipital lobe. Finally, we address whether the cocktail party listening paradigm can further differentiate between them, and we find that it does: selective attention allows greater differentiation between the higher-level processing, and, critically, also surfaces neural correlations with behavioral measures.

4.3. Methods

4.3.1. Participants

MEG data was collected from 22 adults (average age 22.6 yrs, 10 female, 21 right handed) who were native English speakers. The participants gave informed consent and received monetary compensation. All experimental procedures were approved by the Internal Review Board of the University of Maryland, College Park. To ensure that the subjects could satisfactorily perform the arithmetic task, only subjects who self-reported that they had taken at least one college level math course were recruited.

4.3.2. Speech stimuli

Monosyllabic words were synthesized with both male and female speakers using the ReadSpeaker synthesizer (https://www.readspeaker.com, 'James' and 'Kate' voices). The language stimuli consisted of 4-word sentences, and the arithmetic stimuli consisted of 5-word equations. Hereafter, arithmetic words are referred to as 'symbols', arithmetic sentences as 'equations', non-arithmetic words as 'words', and non-arithmetic sentences as 'sentences'. The words and symbols were modified to be of constant durations to allow for separate word, symbol, sentence, and equation rates, so that the neural response to each of these could be separated in the frequency domain. The words and symbols were constructed with fixed durations of 375 ms and 360 ms, respectively, giving a word rate of 2.67 Hz, a symbol rate of 2.78 Hz, a sentence rate of 0.67 Hz, and an equation rate of 0.55 Hz. All the words and symbols were

monosyllabic, and hence the syllabic rate is identical to the word/symbol rate. These rates are quite fast for spoken English, and, though intelligible, can be difficult to follow in the cocktail party conditions. Because neural signals below 0.5 Hz are very noisy, however, it was not deemed appropriate to reduce the rates further; preliminary testing showed that these rates were a suitable compromise between ease of understanding and reasonable neural signal to noise ratio. In addition, the rates were selected such that each trial, made of either 10 equations or 12 sentences, would have the same duration (18 s), allowing for precise frequency resolution at both rates.

The individual words and symbols were shortened by removing silent portions before their beginning and after their end, and then manipulated to have fixed durations, using the overlap-add resynthesis method in Praat (Boersma and Weenick, 2018). The words and symbols were respectively formed into sentences and equations (described below) and were lowpass filtered below 4 kHz using a 3rd order elliptic filter (the air-tube system used to deliver the stimulus has a lowpass transfer function with cutoff approximately 4 kHz). Finally, each stimulus was normalized to have approximately equal perceptual loudness using the MATLAB 'integratedLoudness' function.

The equations were constructed using a limited set of symbols consisting of the word 'is' (denoting '='), three operators ('plus' (+), 'less' (-) and 'times' (×)), and the eleven English monosyllabic numbers ('nil', 'one' through 'six', 'eight' through 'ten', and 'twelve'). The equations themselves consisted of a pair of monosyllabic operands (numbers) joined by an operator, an 'is' statement of equivalence, and a monosyllabic result; the result could be either the first or last symbol in the equation (e.g., 'three plus

two is five' or 'five is three plus two'). The equations were randomly generated with repetitions allowed, in order to roughly balance the occurrences of each number (although smaller numbers are still more frequent since there are more mathematically correct equations using only the smallest numbers). The fact that there were a limited set of symbols and that the same symbol 'is' occurs in every sentence, in either the 2nd or 4th position, are additional regularities, which contribute to additional peaks in the acoustic stimulus spectrum at the first and second harmonic of the equation rate (1.11 Hz and 1.66 Hz) as seen in Fig. 4.1 (and borne out by simulations). Although less than ideal, it is difficult to avoid in a paradigm when restricting to mathematically well-formed equations. Hence, we do not analyze the neural responses at those harmonic frequencies, since their relative contributions from auditory vs. arithmetic processing are not simple to estimate. The sentences were also constructed with two related syntactic structures to be similar to the two equation formats: verb in second position (e.g., 'cats drink warm milk') and verb in third position (e.g., 'head chef bakes pie'), but unlike the 'is' of the arithmetic case, the verb changed with every sentence and there were no analogous harmonic peaks in the sentence case. Deviants were also constructed: deviant equations were properly structured but mathematically incorrect (e.g., 'one plus one is ten'); analogously, deviant sentences were syntactically correct but semantically nonsensical (e.g., 'big boats eat cake'). Cocktail party stimuli were constructed by adding the acoustic waveforms of the sentences and equations in a single audio channel (see Fig. 4.1) and presented diotically (identical for both ears). The speakers were different (male and female), in order to simplify the task of segregating

diotically presented speech. The mixed speech was then normalized to have the same loudness as all the single speaker stimuli using the abovementioned algorithm.

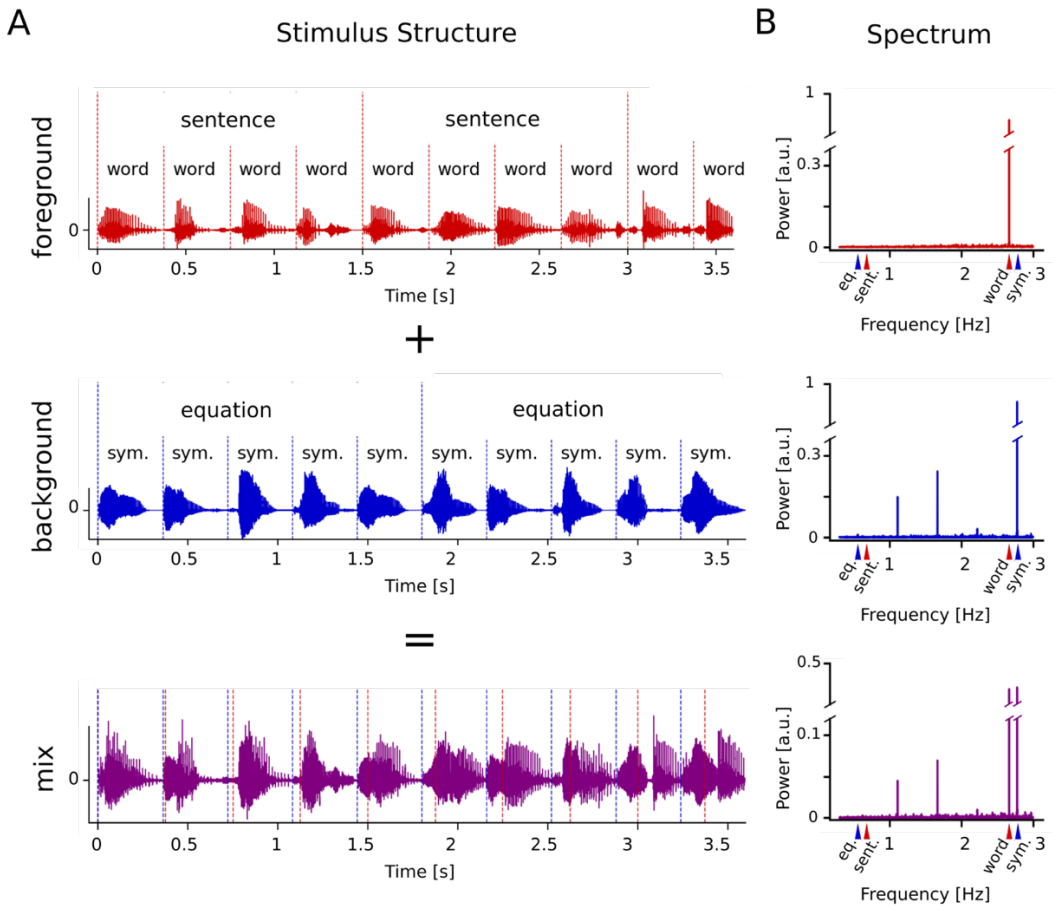

Figure 4.1. Stimulus structure. A. The foreground, background, and mix waveforms for the initial section of the stimulus for a two-speaker attend-language trial. The sentence, equation, word, and symbol structures are shown. The word and symbol rhythms are clearly visible in the waveforms. The mix was presented diotically and is the linear sum of both streams. **B.** The frequency spectrum of the Hilbert envelope of the entire concatenated stimulus for the attend-sentences condition (432 s duration). The sentence (0.67 Hz), equation (0.55 Hz), word (2.67 Hz) and symbol (2.78 Hz) rates are indicated by colored arrows under the x-axis. Clear word and symbol rate peaks are seen in the foreground and background respectively, while the mix spectrum has both peaks. Note that there

are no sentence rate or equation rate peaks in the stimulus spectrum. The appearance of harmonics of the equation rate are consistent with the limited set of math symbols used.

4.3.3. Experimental design

The experiment was conducted in blocks: 4 single speaker blocks (2 × 2: male and female, sentences and equations) were followed by 8 cocktail party blocks (see Table 1). The order of the gender of the speaker was counterbalanced across subjects. Each block consisted of multiple trials: 10 for single speaker and 6 for cocktail party as shown in Table 1. 50% of the blocks had one deviant trial. Each trial consisted of 10 equations or 12 sentences (or both, for cocktail party conditions) and was 18 s in duration for all cases (0.360 s/symbol × 5 symbols/equations × 10 equations = 18 s; 0.375 s/word × 4 words/sentence × 12 sentences = 18 s). In total, the single speaker conditions had 240 sentences and 200 equations, and the cocktail party conditions had 288 sentences and 240 equations in the foreground. Deviant trials had 4 equations or 5 sentences being deviants. At the start of each block, the subject was instructed which stimulus to attend to, and was asked to press a button at the end of each trial to indicate whether a deviant was detected (right button: yes; left button: no). The subjects kept their eyes open, and a screen indicated which voice they should attend to (‘Attend Male’ or ‘Attend Female’) while the stimulus was presented diotically. After each trial, the stimulus was paused, and the screen displayed the text ‘Outlier?’ until the subjects pressed one of the two buttons. There was a 2-second break after the button press, after which the next trial stimulus was presented.

Since the deviant detection task was challenging, especially in the cocktail party case, subjects were asked to practice detecting deviants just before they were placed inside the MEG scanner (2 trials of language, 2 trials of arithmetic, with stimuli not used during the experiment). Most subjects reported that it was easier to follow and detect deviants in the equations compared to the sentences. This might arise for several reasons, e.g., because the equations had a restricted set of simple numbers, or because the repetitive 'is' symbol helped keep track of equation structure.

This experiment was not preregistered. The data is available at https://doi.org/10.13016/xd2i-vyke and the code is available at https://github.com/jpkulasingham/cortical-sentence-equation.

Table 4.1. Experiment Block Structure

Foreground	Background	Speaker Foreground (Background)	Number of trials per block
Equations	-	Male	10
Sentences	-	Male	10
Equations	-	Female	10
Sentences	-	Female	10
Sentences	Equations	Female (Male)	6
Equations	Sentences	Female (Male)	6
Sentences	Equations	Male (Female)	6
Equations	Sentences	Male (Female)	6
Sentences	Equations	Female (Male)	6
Equations	Sentences	Female (Male)	6
Sentences	Equations	Male (Female)	6
Equations	Sentences	Male (Female)	6

The experiment consisted of 4 single speaker blocks followed by 8 cocktail party blocks. Each trial was 18 s in duration and consisted of 10 equations (1.8 s × 10 = 18 s) or 12 sentences (1.5 s × 12 = 18 s). The speaker gender was counterbalanced across subjects (i.e., the order of column 3 was changed).

4.3.4. MEG data acquisition and preprocessing

A 157 axial gradiometer whole head MEG system (Kanazawa Institute of Technology, Nonoichi, Ishikawa, Japan) was used to record MEG data while subjects rested in the supine position in a magnetically shielded room (VAC, Hanau, Germany). The data was recorded at a sampling rate of 2 kHz with an online 500 Hz low pass filter, and a 60 Hz notch filter. Saturating channels were excluded (approximately two channels on average) and the data was denoised using time-shift principal component analysis (de Cheveigné and Simon, 2007) to remove external noise, and sensor noise suppression (de Cheveigné and Simon, 2008) to suppress channel artifacts. All subsequent analyses were performed in mne-python 0.19.2 (Gramfort, 2013; Alexandre Gramfort et al., 2014) and eelbrain 0.33 (Brodbeck et al., 2020). The MEG data was filtered from 0.3–40 Hz using an FIR filter (mne-python 0.19.2 default settings), downsampled to 200 Hz, and independent component analysis was used to remove artifacts such as eye blinks, heartbeats, and muscle movements.

4.3.5. Frequency domain analysis

The complex-valued spectrum of the MEG response for each sensor was computed using the Discrete Fourier Transform (DFT). The preprocessed MEG responses were separated into 4 conditions: attending math or language, in single speaker or cocktail party conditions. The male and female speaker blocks were combined for all analysis. Within each condition, the MEG responses for each trial were concatenated to form

signals of duration 6 minutes for each of the single speaker conditions and 7.2 minutes for each of the cocktail party conditions. The DFT was computed for each sensor in this concatenated response, leading to a frequency resolution of 2.7×10^{-3} Hz for the single speaker conditions and 2.3×10^{-3} Hz for the cocktail party conditions. The amplitudes of the frequency spectra were averaged over all sensors and tested for significant frequency peaks (described in section 4.3.9).

Frequencies of interest were selected corresponding to the equation rate (0.555 Hz), the sentence rate (0.667 Hz), the symbol rate (2.778 Hz), and the word rate (2.667 Hz). Note that the duration of the signals is an exact multiple of both the symbol and the word durations, ensuring that the frequency spectrum contained an exact DFT value at each of these four rates. In addition, the neighboring 5 frequency values (width of ~0.01 Hz) on either side of these key frequencies were also selected to be used in a noise model for statistical tests.

4.3.6. Neural source localization

The head shape of each subject was digitized using a Polhemus 3SPACE FASTRAK system, and head position was measured before and after the experiment using five marker coils. The marker coil locations and the digitized head shape were used to co-register the template FreeSurfer 'fsaverage' brain (Fischl, 2012) using rotation, translation and uniform scaling. A volume source space was formed by dividing the brain volume into a grid of 12 mm sized voxels. This source space was

used to compute an inverse operator using minimum norm estimation (MNE) (Hämäläinen and Ilmoniemi, 1994), with a noise covariance estimated from empty room data. Each sensor's response was concatenated over all trials in each condition, and the Fourier transform was used to compute the complex-valued frequency spectrum—with both amplitude and phase—at that sensor. The values of these spectra at each of the 44 selected frequencies (4 frequencies of interest with ten sidebands each) are complex-valued sensor distributions, which were source-localized independently using MNE onto the volume source space, giving complex-valued source activations (Simon and Wang, 2005). The amplitudes of these complex-valued source activations were used for subsequent analysis. Finally, the sideband source distributions were averaged together to form the noise model.

4.3.7. Temporal Response Functions (TRFs)

The preprocessed single-trial MEG responses in each of the four conditions (excluding deviant trials) were source-localized in the time domain using MNE, similar to the method described above in the frequency domain. The MEG signals were further lowpassed below 10 Hz using an FIR filter (default settings in mne python) and downsampled to 100 Hz for the TRF analysis. These responses were then used along with representations of the stimulus to estimate TRFs. The linear TRF model for P predictors (stimulus representations) is given by

$$y(t) = \sum_{p=1}^{P} \sum_{d} \tau^{(p)}(d)\, x^{(p)}(t-d) \; + n(t) \tag{4.1}$$

where $y(t)$ is the response at a neural source at time t, $x^{(p)}(t-d)$ is the time shifted p^{th} predictor (e.g., speech envelope, word onsets, etc., as explained below) with time lag of d, $\tau^{(p)}(d)$ is the value of the TRF corresponding to the p^{th} predictor at lag d, and $n(t)$ is the residual noise. The TRF estimates the impulse response of the neural system for that predictor, and can be interpreted as the average time-locked response to continuous stimuli (Lalor and Foxe, 2010). For this analysis, several predictors were used to estimate TRFs at each neural source using the boosting algorithm (David et al., 2007), as implemented in eelbrain, thereby separating the neural response to different features. The boosting algorithm may result in overly sparse TRFs, and hence an overlapping basis of 30 ms Hamming windows (with 10 ms spacing) was used in order to allow smoothly varying responses. For the volume source space, the TRF at each voxel for a particular predictor is a vector that varies over the time lags, representing the amplitude and direction of the current dipole activity.

The stimulus was transformed into two types of representations that were used for TRF analysis: acoustic envelopes and rhythmic word/symbol or sentence/equation onsets. Although we were primarily interested in responses to sentences and equations, a linear model with only sentence/equation onsets would be disadvantaged by the fact that these representations are highly correlated with the acoustics. Hence by jointly estimating the acoustic envelope and word onset TRFs in the model, the lower-level acoustic responses are automatically separated, allowing the dominantly higher-level

processing to emerge in the sentence/equation TRFs. The acoustic envelope was constructed using the 1-40 Hz bandpassed Hilbert envelope of the audio signal (FIR filter used above). The onset representations were formed by placing impulses at the regular intervals corresponding to the onset of the corresponding linguistic unit. The four onset responses were: impulses at 375 ms spacing for word onsets, 360 ms for symbol onsets, 1500 ms for sentence onsets, and 1800 ms for equation onsets. Values at all other time points in these onset representations were set to zero. In order to separate out responses to stimulus onset and offset, the first and last sentences were assigned separate onset predictors, which were not analyzed further except to note that their TRFs showed strong and sustained onset and offset responses, respectively. The remaining (middle) sentences' onsets were combined into one predictor that was used for further TRF analysis. The same procedure was followed for the equation onset predictors.

For each of the two single-speaker conditions, five predictors were used in the TRF model: the corresponding three sentence/equation onsets (just described), word/symbol onsets and the acoustic envelope. For each of the two cocktail conditions, ten predictors were used in the TRF model: the abovementioned five predictors, for each of the foreground and the background stimuli. The predictors were fit in the TRF model jointly, without any preference given to one of them over another.

The TRF for the speech envelope and the word/symbol onsets were estimated for time lags of 0-350 ms in order to limit the TRF duration to before the onset of the next word (at 375 ms) or symbol (at 360 ms). The sentence and equation TRFs were

estimated starting from 350 ms to avoid onset responses, as well as lagged responses to the previous sentence. The sentence TRF was estimated until 1850 ms (350 ms past the end of the sentence) and the equation TRF was estimated until 2150 ms (350 ms past the end of the equation), in order to detect lagged responses. These sentence and equation TRFs were used to further analyze high level arithmetic and language processing.

4.3.8. Decoder analysis

All decoding analyses were performed using scikit-learn (Pedregosa et al., 2011) and mne-python software. To investigate the temporal dynamics of responses, linear classifiers were trained on the MEG sensor space signals bandpassed 0.3-10 Hz at 200 Hz sampling frequency. Decoders were trained directly on the sensor space signals, since the linear transformation to source space cannot increase the information already present in the MEG sensor signals. The matrix of observations $\mathbf{X} \in \mathbb{R}^{N \times M}$, for N samples and M sensors in each sample, was used to predict the vector of labels $\mathbf{y} \in \{0, 1\}^N$ at each time point of sentences or equations. The labels correspond to the two attention conditions: attend-equations or attend-sentences. The decoders were trained in the single speaker conditions on time points from 0 to 1500 ms for both 1500 ms long sentences and 1800 ms long equations. Therefore, the decoder at each time point learns to predict the attended stimulus type (equations or sentences) using the MEG sensor topography at that time point.

In a similar manner, the operator type in the arithmetic condition was also decoded from the MEG sensor topographies at each time point, in the 720 ms time window of each equation that contained the operator and its subsequent operand. 3 decoders were trained for the 3 comparisons ('plus' vs. 'less', 'less' vs. 'times' and 'plus' vs. 'times').

To further investigate the patterns of cortical activity, linear classifiers were trained on the source localized MEG responses at each voxel, with $\mathbf{X} \in \mathbb{R}^{N \times T}$ for N samples and T time points in each sample. The response dynamics of the entire sentence/equation may not be suitable for decoding: since the equations are comprised of five symbols, while the sentences of four words, this might lead to decoding the equations vs. sentences based on whether there were five vs. four auditory responses to acoustic onsets. To minimize this confound, two types of classifiers were used based on responses to only one word/symbol (and hence with only one acoustic onset). 1) Decoding based on first words: The first symbol of each equation and first word of each sentence was used as the sample, with a label denoting attend equations or attend sentences conditions. 2) Decoding based on last words: The last symbol or word was used. Words of duration 375 ms were downsampled to match the duration of the symbols (360 ms), in order to have equal length training samples. This method was used separately for both the single speaker and the cocktail party conditions. The decoder at each voxel learns to predict the attended stimulus type (equations or sentences) using the temporal dynamics of the response at that voxel.

Finally, the effect of attention was investigated using two sets of classifiers for equations and sentences at each voxel. For the attend-equations classifier, the cocktail party trials were separated into samples at the twelve equation boundaries, and the labels were marked as '1' when math was attended to and '0' when not. The time duration T was 0-1800 ms (entire equation). For the attend-sentences classifier, the cocktail party trials were separated into samples at the ten sentence boundaries and the labels were '1' when attending to sentences and '0' otherwise. The time duration T was 0-1500 ms (entire sentence). Therefore, the attend-equations decoder at each voxel learns to predict whether the equation stimulus was attended to using the temporal dynamics of the response to the equation at that voxel (and similarly for the attend-sentences decoder).

In summary, the decoders at each time point reveal the dynamics of decoding attention to equations vs. sentences from MEG sensor topographies, and the decoders at each voxel reveal the ability to decode arithmetic and language processing in specific cortical areas. The trained classifiers were tested on a separate set and the score of the decoder was computed. Logistic regression classifiers were used, with 5-fold cross-validation, within-subject for all the trials. The area under the receiver operating characteristic curve (*AUC*) was used to quantify the performance of the classifiers.

4.3.9. Statistical analysis

Two types of nonparametric permutation tests were performed across subjects to control for multiple comparisons: single threshold max-t tests for the amplitude spectra,

and cluster based permutation tests for the source localized responses. For the former case, the amplitude spectra for each condition were averaged across sensors, and permutation tests were used to detect significant peaks across subjects (n=22). Each frequency value in the spectrum from 0.3 to 3 Hz was tested for a significant increase over the average of the neighboring 5 values on either side using 10000 permutations and the single threshold max-t method (Nichols and Holmes, 2002) to adjust for multiple comparisons. In brief, a null distribution of max-t values was calculated using the maximum t-values obtained across all frequencies, for each permutation. Any t-value in the observed frequency spectra (denoted by t_{obs}) that exceeds the 95th percentile of the max-t null distribution was deemed significant. For these tests, we report the p-values and the t_{obs} values, and deliberately omit the degrees of freedom to avoid direct comparison between the two, since the p-values are derived entirely from the permutation distribution of max-t values and not from the t-distribution. Correlation tests were also performed to investigate associations between different responses (e.g., sentence rate vs. equation rate) within each subject. Pearson correlation tests with Holm-Bonferroni correction were used on the responses at the frequencies of interest, after subtracting the average of the five neighboring bins on either side.

Cluster based permutation tests were performed for the source-localized responses. The source distributions for each individual were mapped onto the FreeSurfer 'fsaverage' brain, in order to facilitate group statistics. To account for individual variability and mislocalization during this mapping, the distributions were spatially smoothed using a Gaussian window with a standard deviation of 12 mm for all

statistical tests. The source localized frequency responses were tested for a significant increase over the corresponding noise model formed by averaging the source localized responses of the five neighboring frequencies on either side. Nonparametric permutation tests (Nichols and Holmes, 2002) and Threshold Free Cluster Enhancement (TFCE) (Smith and Nichols, 2009) were performed to compare the response against the noise and to control for multiple comparisons. A detailed explanation of this method can be found in Brodbeck et al. (2018). Briefly, a test statistic (in this case, paired samples t-statistics between true responses and noise models) is computed for the true data and 10000 random permutations of the data labels. The TFCE algorithm is applied to these statistics, in order to enhance continuous clusters of large values, and a distribution consisting of the maximum TFCE value for each permutation is formed. Any value in the original TFCE map that exceeds the 95th percentile is considered significant at the 5% significance level. In all subsequent results, the minimum p-value and the maximum or minimum t-value across voxels is reported as p_{min}, t_{max} or t_{min} respectively. Note that the p_{min} is derived from the permutation distribution and cannot be derived directly from t_{max} or t_{min} using the t-distribution (degrees of freedom are also omitted due to this reason). Lateralization tests were performed by testing each voxel in the left hemisphere with the corresponding voxel in the right, using permutation tests and TFCE with paired samples t-statistics. For the attend math conditions, equations were separated by operator type ('plus' (+), 'less' (-) or 'times' (×)) to test for specific responses to each operator. No

significant differences were found between the source localized responses to each operator.

To test for significant effects and interactions, repeated measures ANOVAs were performed on the source localized responses at each frequency of interest after subtracting the corresponding noise model. Nonparametric permutation tests with TFCE were used to correct for multiple comparisons, similar to the method described above. In brief, a repeated-measures ANOVA is performed at each voxel, and then, for each effect or interaction, the voxel-wise F-values from this ANOVA are passed into the TFCE algorithm, followed by permutation tests as described earlier. This method detects significant clusters in source space for each significant effect. Note that the maximum F-value in the original map within a cluster (F_{max}) and the p-value of the cluster are reported (and degrees of freedom omitted), for the same reasons as those explained in the previous paragraph (i.e., p-values are derived from the permutation distribution and not the F-distribution).

Several types of repeated measures ANOVAs were performed using the abovementioned method. In the single speaker case, a 2 × 2 ANOVA with factors stimulus ('language' for words/sentences or 'math' for symbols/equations) and frequency ('low' for sentence/equation and 'high' for word/symbol) was performed. For the cocktail party case, a 2 × 2 × 2 ANOVA with the added factor of attention (attended or unattended) was performed. In addition, two further ANOVAs were performed to investigate hemispheric effects, using an additional factor of hemisphere

for both the single speaker (2 × 2 × 2 ANOVA) and the cocktail party (2 × 2 × 2 × 2 ANOVA) conditions.

To investigate significant ANOVA effects further, post-hoc t-tests across subjects were performed on the responses averaged across voxels within the relevant significant cluster. For this scalar t-test, a Holm-Bonferroni correction was applied to correct for multiple comparisons. For these tests, the t-values with degrees of freedom, corrected p-values and Cohen's d effect sizes are reported.

Behavioral responses for the deviant detection task were classified as either correct or incorrect, and the number of correct responses for each subject was correlated with the source localized response power of that subject. The noise model for each frequency of interest was subtracted from the response power before correlating with behavior. Nonparametric permutation tests with TFCE were used in a manner similar to that given above. The only difference was that the statistic used for comparison was the Pearson correlation coefficient between the two variables (behavior and response power), and the maximum correlation coefficient across voxels is reported as r_{max}.

The TRFs were tested for significance using vector tests based on Hotelling's T^2 statistic (Mardia, 1975). Since the TRFs consist of time-varying vectors, this method tests consistent vector directions across all subjects at each time point and each voxel. The Hotelling's T^2 statistic was used with non-parametric permutation tests and TFCE as described above, with the added dimension of time, and the maximum T^2 statistic across voxels is reported as T^2_{max}. This statistic is more suitable than a t-statistic based on the amplitude of the TRF vectors, since activation from distinct neural processes

may have overlapping localizations (due to the limited resolution of MEG), but different current directions.

Finally, the decoders were tested across subjects for a significant increase in decoding ability above chance ($AUC = 0.5$) at each time point or at each voxel. Multiple comparisons were controlled for using permutation tests and TFCE, similar to the above cases, with AUC as the test statistic.

4.4. Results

4.4.1. Behavioral results

After each trial, subjects indicated whether that trial contained a deviant by pressing a button. The single speaker conditions had higher deviant detection accuracies (equations: mean = 89.5%, SD = 10.7%; sentences: mean = 73.4%, SD = 13.8%) than the cocktail party conditions (equations: mean = 79.9%, SD = 13.3%; sentences: mean = 61%, SD = 19.4%). Subjects reported that the equations were perceptually easier to follow than the sentences, consistent with the fact that the equations were formed using a smaller set of monosyllabic numbers to preserve the symbol rates. The presence of ‘is’ in each equation may have also contributed to subjects tracking equation boundaries.

4.4.2. Frequency domain analysis

The response power spectrum was averaged over all sensors and a permutation test with the max-t method was performed to check whether the power at each frequency of interest was significantly larger than the average of the neighboring five frequency bins on either side (see Fig. 4.2 A, B) across subjects (n = 22). For the language single speaker condition, the sentence rate (0.67 Hz, t_{obs} = 7.25, p < 0.001, Note: degrees of freedom not shown since p-values are derived from the permutation test, see Methods 4.3.9), its first harmonic (1.33 Hz, t_{obs} = 6.11, p = 0.0023), and the word rate (2.67 Hz, t_{obs} = 12.98, p < 0.001) were significant (one tailed permutation test of difference of amplitudes with max-t method). Similarly, for the math single speaker condition, the symbol rate (2.78 Hz, t_{obs} = 12.39, p < 0.001) and the equation rate (0.55 Hz, t_{obs} = 6.29, p = 0.0017) were significant. In this condition, the 1st and 2nd harmonics of the equation rate were also significant (t_{obs} = 7.28, p < 0.001 at 1.11Hz; t_{obs} = 7.77, p < 0.001 at 1.67 Hz). Thus, in both conditions, the responses track the corresponding sentence or equation rhythms that are not explicitly present in the acoustic signal. The harmonic peak (1.33 Hz) in the language condition is consistent with phrase tracking (Ding et al., 2016), and the harmonics in the arithmetic condition (1.11 Hz, 1.66 Hz) are consistent with auditory processing of acoustic properties of the stimulus associated with the limited number of mathematical symbols employed (see Methods 4.3.2), or higher-order processing, or both. Correlation tests within subjects, with Holm-Bonferroni correction, were performed on relevant pairs of responses (after subtracting the neighboring bins). Sentence rate responses were significantly correlated with equation

rate responses (Pearson's $r = 0.576$, $p = 0.015$). Word rate responses were significantly correlated with symbol rate responses ($r = 0.681$, $p = 0.001$). Since such correlations may arise from fluctuating degree of task engagement, or variable neural signal to noise ratio across subjects, they were not analyzed further. There were no significant sentence vs. word ($r = 0.067$, $p > 0.99$) or equation vs. symbol ($r = 0.001$, $p > 0.99$) response correlations.

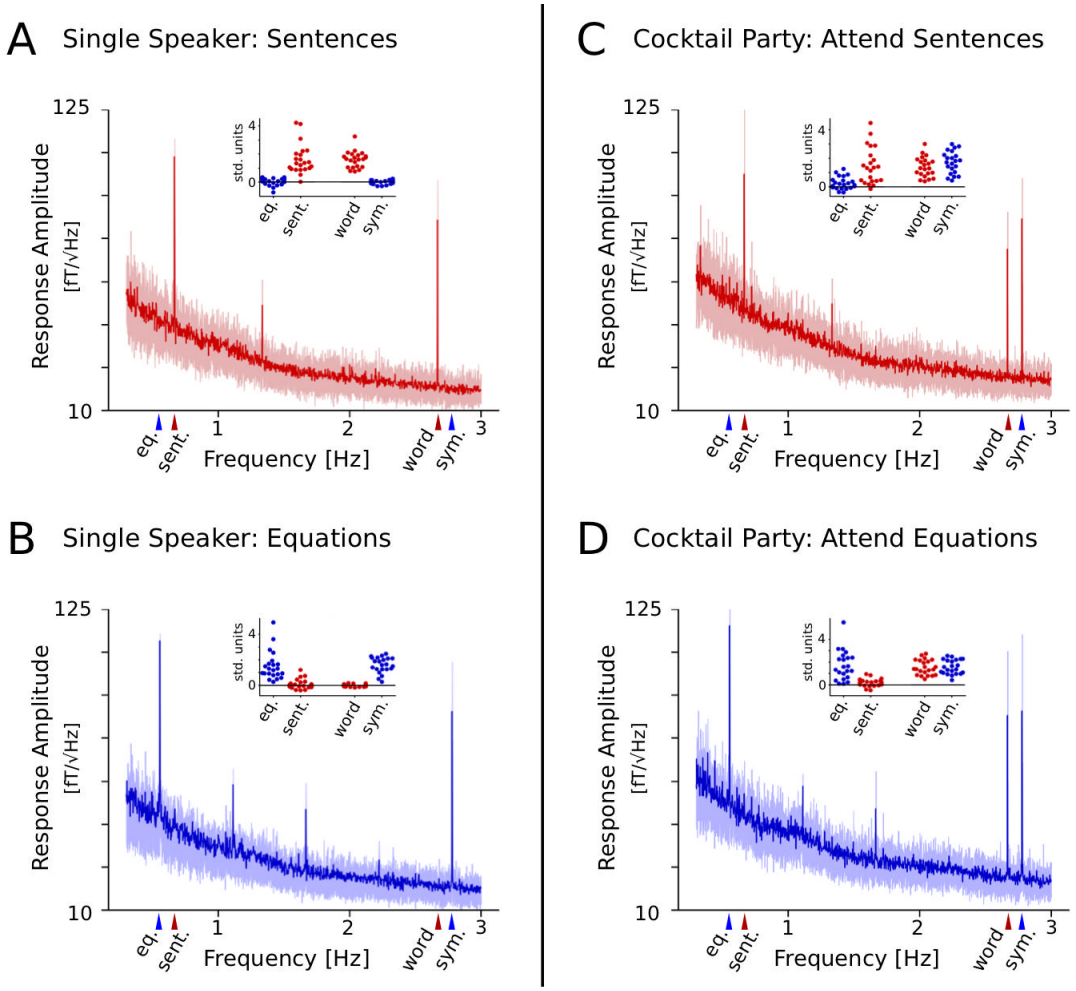

Figure 4.2. Neural response spectrum. The MEG response spectrum as a function of frequency for the four conditions. The amplitude spectrum, averaged over sensors and subjects, is shown with light shaded regions denoting the 1st-3rd quartile range across subjects. Clear peaks are seen at the

sentence, equation, word, and symbol rates (indicated by the arrows under the x-axis). These responses were compared against neighboring bins (of width ~0.01 Hz, not visible here) for statistical tests. Insets show the average responses at the four frequencies of interest for each subject, after subtracting the neighboring bins. The scale for the insets is standardized within each condition, but with 0 indicating the baseline average activity of the neighboring bins. For the single speaker conditions, peaks appear only at the rates corresponding to the presented stimulus. For the cocktail party conditions, peaks appear at the symbol and word rates regardless of attention, while sentence and equation peaks only appear during the attended condition. There are no analogous sentence or equation peaks during the opposite attention condition.

For the attend-sentences cocktail party condition, only the (attended) word, (unattended) symbol and (attended) sentence rate responses were significant (one tailed permutation test of difference of amplitudes with max-t method; $t_{obs} = 9.29$, $p < 0.001$; $t_{obs} = 10.59$, $p < 0.001$; $t_{obs} = 5.46$, $p = 0.0176$ respectively) as shown in Fig 4.2 C, D. The (unattended) equation rate response was not significant ($t = 2.99$, $p > 0.99$). On the other hand, for the attend-equations cocktail party condition, the (unattended) word, (attended) symbol and (attended) equation rate responses were significant ($t_{obs} = 10.86$, $p < 0.001$; $t_{obs} = 11.64$, $p < 0.001$; $t_{obs} = 6.07$, $p = 0.005$ respectively), while the (unattended) sentence rate response was not significant ($t_{obs} = 2.73$ $p > 0.99$). Responses at the 1st and 2nd harmonics of the equation rate were also significant in the attend-equations condition (1.11 Hz, $t_{obs} = 5.31$, $p = 0.027$; 1.67 Hz, $t_{obs} = 5.09$, $p = 0.04$). Correlation tests within subjects were performed, similar to the single speaker case, on all responses except the non-significant unattended sentence and equation rates. Once

again, attended sentence rate responses were significantly correlated with attended equation rate responses ($r = 0.68$, $p = 0.0023$). Word rate responses were significantly correlated with symbol rate responses for both attended ($r = 0.69$, $p = 0.0021$) and unattended cases ($r = 0.83$, $p < 0.001$). Other correlations were not significant (attended sentence vs. attended word: $r = 0.12$, $p > 0.99$; attended sentence vs. unattended word: $r = 0.07$, $p = 0.74$; attended equation vs. attended symbol: $r = 0.49$, $p = 0.083$; attended equation vs. unattended symbol: $r = 0.33$, $p = 0.4$).

Since the word and symbol rates are present in the acoustics for both conditions, the neural responses at these rates could merely reflect acoustic processing. However, the fact that the sentence and equation rates are significant only in the corresponding attention condition suggests that these responses may dominantly reflect attention-selective high-level processes. This agrees with prior studies showing a similar effect for language (Ding et al., 2018). Here we show that this effect occurs even for arithmetic equations. However, arithmetic equations are also sentences, so it is unclear from this result alone if the equation peak reflects merely tracking of sentential structure and not arithmetic processing. To investigate this, we used volume source localization on the responses at the relevant frequencies to determine the cortical distribution of these responses.

The responses at the 4 frequencies of interest (word, symbol, sentence and equation rates) were source-localized using the Fourier transform sensor topographies at these frequencies (see Methods 4.3.6). The amplitudes of the resulting complex-valued volume source distributions were used for all subsequent analysis. For each frequency

of interest, the source amplitudes in the neighboring five bins on either side were calculated using the same source model, and averaged together as an estimate of the background noise. The response distributions for each of these frequencies were tested for a significant increase over the noise estimate using nonparametric permutation tests with paired sample t-statistics and TFCE. For both single speaker conditions, the corresponding word or symbol responses were significant (t_{max} = 12.85, p_{min} < 0.001, and t_{max} = 12.77, p_{min} < 0.001, respectively) in the regions shown in Fig. 4.3 A, B, with the average response being strongest in bilateral auditory cortex. The word and symbol rate responses were not significantly different (t_{min} = -2.11, t_{max} = 3.63, p > 0.08), consistent with low level auditory processing. The corresponding sentence or equation responses were also significant (t_{max} = 9.92, p_{min} < 0.001, and t_{max} = 7.68, p_{min} < 0.001, respectively). The source distribution for sentence responses was predominantly in left auditory cortex and temporal lobe, whereas for equations the response was distributed over areas of bilateral temporal, parietal, and occipital lobes. Despite these visually distinct patterns, the two responses were not significantly different (t_{min} = -3.00, t_{max} = 3.36, p > 0.12), perhaps because large portions of the brain show activity synchronized to the rhythm. Both sentence and equation responses were significantly left lateralized in temporal (t_{max} = 6.69, p_{min} < 0.001) and parietal (t_{max} = 3.9, p_{min} = 0.009) areas respectively. No significant differences were seen in the responses at the equation rate when separated according to operator type ('+' vs. '-': t_{min} = -2.98, t_{max} = 1.64, p > 0.34; '-' vs. '×': t_{min} = -2.08, t_{max} = 3.21, p > 0.39; '×' vs. '+': t_{min} = -2.26, t_{max} = 3.01, p > 0.31).

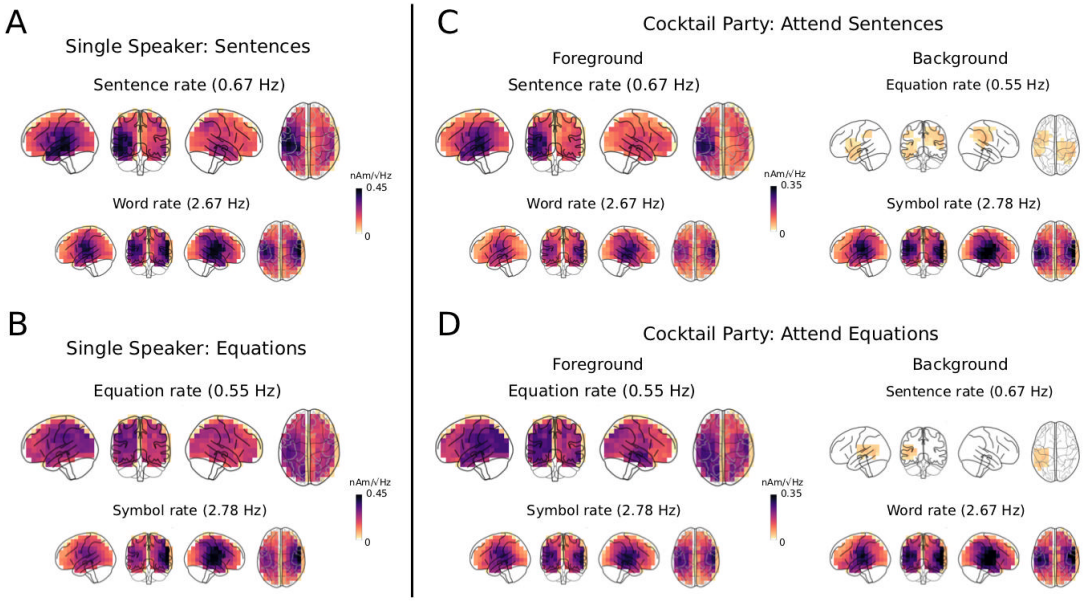

Figure 4.3. Source localized responses at each frequency of interest. The source localized responses at critical frequencies, averaged over subjects and masked by significant increase over the noise model, are shown. Color scales are normalized within each condition in order to more clearly show the spatial patterns. The word and symbol rate responses are maximal in bilateral auditory cortical areas, while the sentence rate response is maximal in the left temporal lobe. The equation rate responses localize to bilateral parietal, temporal, and occipital areas, albeit with increased left hemispheric activity. Although the background sentence and equation rates also show significant activity, the amplitude of these responses are much smaller than the responses at the corresponding attended rates.

For the cocktail party conditions, similar results were obtained for both word and symbol rate responses (attend sentences: word rate: $t_{max} = 11.9$, $p_{min} < 0.001$, symbol rate: $t_{max} = 12.8$, $p_{min} < 0.001$; attend equations: word rate: $t_{max} = 11.1$, $p_{min} < 0.001$, symbol rate: $t_{max} = 11.01$, $p_{min} < 0.001$). The response was predominantly in bilateral

auditory cortices as shown in Fig. 4.3 C, D, and the symbol and word rates were not significantly different (t_{min} = -4.31, t_{max} = 2.33, $p > 0.16$). The attended sentence or equation rate responses were significant (t_{max} = 6.78, $p_{min} < 0.001$, and t_{max} = 7.87, $p_{min} < 0.001$, respectively) and the localization was similar to the single speaker case, albeit more bilateral for the equation rate response. Indeed, the sentence rate response was significantly left lateralized (t_{max} = 5.36, $p_{min} < 0.001$), similar to the single speaker case, but the equation rate response was not (t_{max} = 2.97, $p > 0.067$). However, the spatial distribution of the equation rate response was larger in the left hemisphere (see Fig. 4.3); indeed, the source localization of attended sentence responses and attended equation responses were significantly different (t_{min} = -4.77, t_{max} = 2.39, p_{min} = 0.013), with more equation rate responses in the right hemisphere. This indicates that the equation rate response does not originate from the same cortical regions that give rise to the sentence rate response and that the selective attention task is better able to separate these responses. Perhaps surprisingly, the unattended sentence and equation rates were also significant (t_{max} = 4.02, p_{min} = 0.005, and t_{max} = 5.31, $p_{min} < 0.001$, respectively) in small clusters, even though such peaks do not appear in the frequency spectrum averaged across all sensors (Fig. 4.2). Note however, that some individuals did show small peaks at these rates even in the average spectrum (see points above zero for unattended rates in the insets of Fig. 4.2 C, D).

A repeated measures ANOVA was performed for the single speaker case on the abovementioned source space distributions (as shown in Fig. 4.3) for each frequency of interest. The 2 × 2 ANOVA consisted of factors stimulus (‘language’ for

word/sentence or ‘math’ for symbol/equation) and specific frequency (‘high’ for word/symbol or ‘low’ for sentence/equation). The ANOVA was performed on the response at each voxel and cluster-based permutation tests with TFCE were used to correct for multiple comparisons (See Methods 4.3.9 for choice of reported statistics). The interaction of stimulus × frequency was not significant (F_{max} = 10.38, p = 0.149, but see below for an interaction effect in an ANOVA with a factor of hemisphere). A significant main effect of frequency (F_{max} = 18.63, p = 0.006) was found in a right auditory cluster and a significant main effect of stimulus type (F_{max} = 21.67, p = 0.003) was found in the left auditory/temporal area. Post-hoc t-tests across subjects were performed on the responses averaged across voxels within the significant clusters for each effect; p-values were obtained from the t-distribution and then corrected for multiple comparisons using the Holm-Bonferroni method. These tests revealed that the main effect of stimulus was due to a significant increase in both the sentence over the equation responses (t(21) = 2.96, p = 0.037, Cohen’s d = 0.54) and the word over the symbol responses (t(21) = 2.85, p = 0.038, Cohen’s d = 0.52) in the left auditory/temporal cluster, consistent with increased left temporal activity for language over arithmetic. The main effect of frequency was due to a significant increase in both the word over the sentence responses (t(21)=3.67, p = 0.01, Cohen’s d=1.16), and the symbol over the equation responses (t(21)=3.15, p = 0.028, Cohen’s d=0.97) in the right auditory cluster.

For the cocktail party case, a similar repeated measures ANOVA was performed, but with an additional factor of attention (attended or unattended) leading to a 2 × 2 × 2 design. A significant 3-way interaction of stimulus × attention × frequency was found in a right parietal cluster (F_{max} = 15.18, p = 0.024). Post-hoc t-tests across subjects with Holm-Bonferroni correction were performed on the responses averaged across voxels within this cluster. These revealed a significant increase in the equation responses compared to the sentence responses when attended ($t(21)$ = 3.71, p = 0.0103, Cohen's d = 0.82), but no significant difference when unattended ($t(21)$ = 2.27, p = 0.09, Cohen's d = 0.65). There was also no significant difference between word and symbol responses both when attended ($t(21)$=-0.32, p = 0.75, Cohen's d = -0.06) and unattended ($t(21)$=-0.69, p = 0.99, Cohen's d = -0.09). This is consistent with increased responses to equations in right parietal areas only when attended. In addition to this 3-way interaction, several 2-way interactions and main effects were also detected but were not analyzed further.

Finally, two further ANOVAs were performed with an additional factor of hemisphere for both the single speaker (2 × 2 × 2 ANOVA) and cocktail party (2 × 2 × 2 × 2 ANOVA). For the single speaker case, the 3-way interaction was significant (stimulus × frequency × hemisphere: F_{max} = 18.55, p = 0.016) in superior parietal voxels. For the cocktail party case, the 4-way interaction was not significant (attention × frequency × stimulus type × hemisphere: F_{max} = 8.31, p=0.115). However, two 3-way interactions involving hemisphere were significant (attention × frequency × hemisphere: F_{max} = 13.71, p = 0.031, frequency × stimulus × hemisphere: F_{max} = 18.12,

p =0.017) in temporal voxels. Other effects involving hemisphere were also found to be significant (frequency × hemisphere: F_{max} = 41.75, p < 0.001, main effect of hemisphere: F_{max} = 17.1, p = 0.021), as well as several other effects not involving hemisphere. These effects were not analyzed further, but they indicate that the effects of attention, stimulus and frequency depend significantly on the hemisphere, as already suggested by the lateralized clusters found in the simpler ANOVAs described earlier.

In summary, the ANOVA analysis indicates that, in the single speaker case, low-level responses (word/symbol) are significantly stronger than the higher-level responses (sentence/equation) in right auditory areas and that the language responses (sentence/word) are significantly stronger than the arithmetic responses (equation/symbol) in left auditory/temporal areas. Critically, the ANOVA results for cocktail party indicate that the equation responses are significantly larger than the sentence responses in right parietal areas but only when attended to. ANOVAs also indicate that these effects depend on hemisphere as already suggested by the previous pairwise comparisons.

4.4.3. Behavioral correlations

Behavioral performance was correlated with source localized neural responses using non-parametric permutation tests with TFCE, with Pearson correlation as the test statistic. Deviant detection performance for sentences in the single speaker condition was significantly correlated with the sentence rate neural response (p_{min} = 0.02, maximum correlation in significant regions r_{max} = 0.62) as shown in Fig. 4.4. However,

detection of equation deviants in the single speaker condition was not significantly correlated with the equation rate neural response; this may be related to the fact that performance in the single speaker arithmetic condition was at ceiling for several participants. The performance when detecting sentence deviants in the cocktail party conditions was correlated with the attended sentence rate response (r_{max} = 0.62, p_{min} = 0.015), attended word rate response (r_{max} = 0.64, p_{min} = 0.03) as well as the unattended symbol rate response (r_{max} = 0.79, p_{min} = 0.001). The performance when detecting equation deviants in the cocktail party condition was correlated with the attended equation rate response (r_{max} = 0.6, p_{min} = 0.02) and the unattended word rate response (r_{max} = 0.74, p_{min} = 0.04). It was unexpected that the unattended word and symbol rate responses were significantly correlated with behavior, and possible explanations are discussed in section 4.5.3. Critically, however, sentence and equation rate responses were correlated with behavior only when attended.

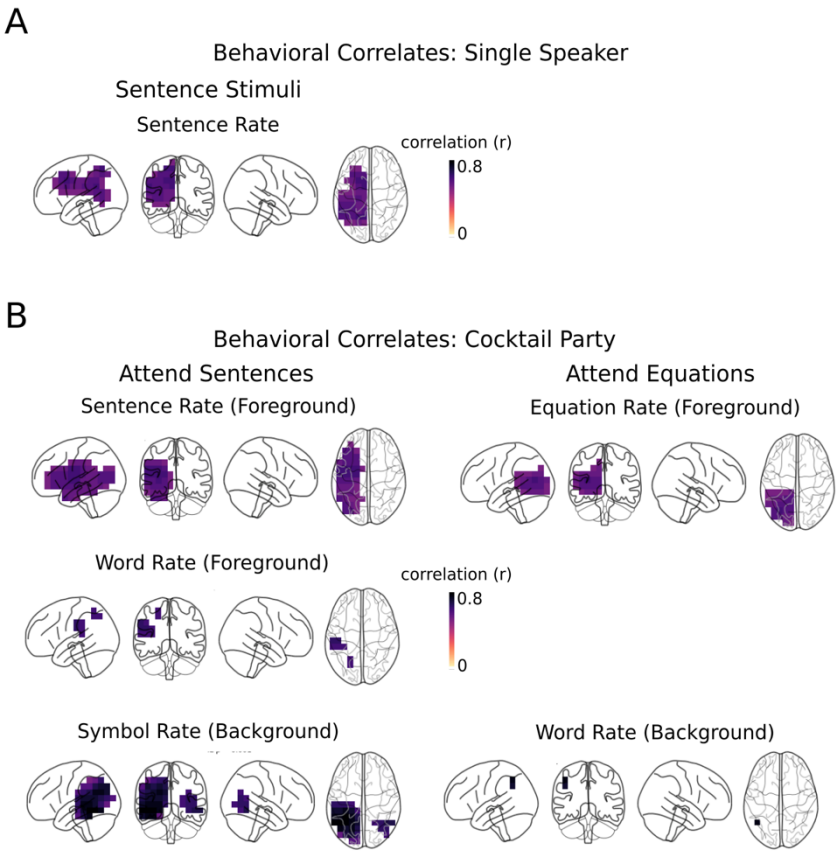

Figure 4.4. Neural response correlations with behavior. The source localized responses at the frequencies of interest were correlated with the corresponding deviant detection performance, across subjects. The areas of significant correlation are plotted here (same color scale for all plots). Sentence and equation rate responses are significantly correlated with behavior only if attended, while both attended and unattended word rate responses are significantly correlated with behavior. The sentence rate response is significantly correlated over regions in left temporal, parietal, and frontal areas, while significant correlation for the equation rate response is seen in left parietal and occipital regions.

4.4.4. TRF analysis

TRF analysis was performed using source localized MEG time signals for each condition after excluding the deviant trials (details in Methods 4.3.7). TRFs were simultaneously obtained for responses to the acoustic envelopes, word/symbol onsets and sentence/equation onsets. Although stimuli with fixed and rhythmic word, symbol, sentence, and equation onsets might lend itself to an evoked response analysis, the fact that the words (or symbols) are only separated by 375 ms (or 360 ms) may lead to high-level late responses overlapping with early auditory responses to the next word (or symbol). In contrast, computing simultaneous TRFs to envelopes and word/symbol onsets in the same model as TRFs to equation/sentence onsets regresses out auditory responses from higher-level responses, providing cleaner TRFs for sentences and equations. The obtained envelope and word/symbol TRFs were not used for further analysis, since they were dominated by acoustic responses that have been well-studied in other investigations (Brodbeck et al., 2018a, 2018b). The volume source localized TRFs are time-varying vectors at each voxel. Activity of nearby neural populations may overlap, even if the activity is driven by different processes, due to the limited spatial resolution of MEG. However, these effects may have different current directions due to the anatomy of the cortical surface. Therefore, a test for consistent vector directions, using Hotelling's T^2 statistic and permutation tests with TFCE, was used to detect group differences in the direction of current flow (see Methods 4.3.9).

The sentence and equation TRFs showed significance over several time intervals and many voxels over the duration of the TRF (T^2_{max} = 7.66, p_{min} < 0.001, and T^2_{max} = 5.12, p_{min} < 0.001, respectively, see Fig 4.5). The TRFs were computed starting from 350 ms after the sentence onset to 350 ms after the end of the sentence, but because of the fixed-period presentation rate without any breaks between sentences in a trial, the TRFs from 0-350 ms are identical to the last 350 ms. The large peak at the end (and beginning) of each TRF may either arise from processing of the completion of the sentence/equation, or from preparation (or auditory) processing of the new sentence sentence/equation, or both. This peak occurs around 60-180 ms after the start of the new sentence/equation, in the typical latency range of early auditory processing. However, spatial distributions of the peak in the equation TRFs seem to indicate patterns that are not consistent with purely auditory processing, especially for the cocktail party condition (described below). Additionally, significant activity is seen throughout the duration of the sentence/equation that is not tied to word/symbol onsets, indicating that lower-level auditory responses have been successfully regressed out. Therefore, the large peaks at the end plausibly reflect processing of the completion of the sentence/equation (with a latency of 420-530 ms after the last word/symbol). The sentence TRF peaks were significant predominantly in the left temporal lobe, while the equation TRF peaks were significant in bilateral temporal, parietal, and motor areas.

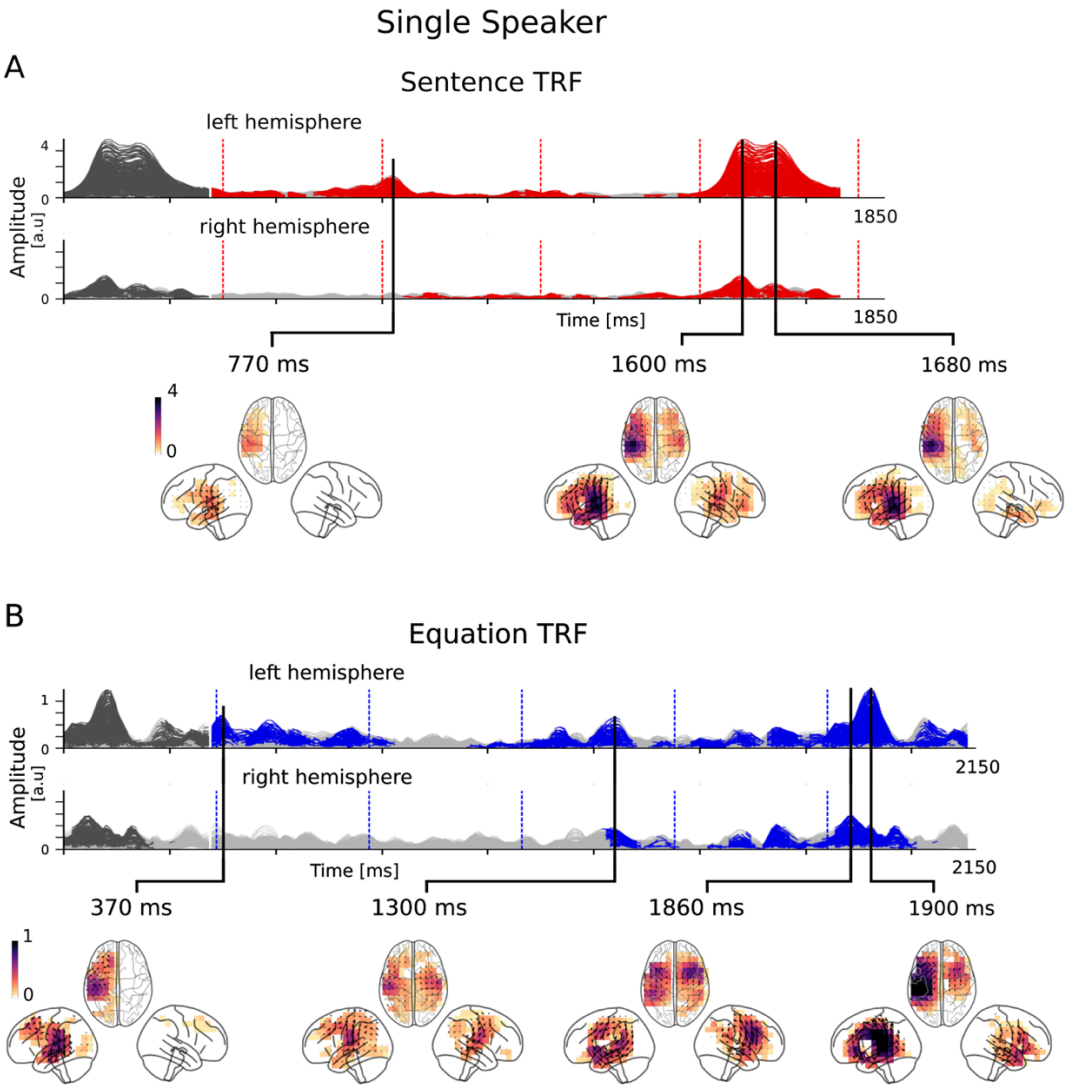

Figure 4.5. TRFs in the single speaker conditions. Overlay plots of the amplitude of the TRF vectors for each voxel, averaged over subjects. For each TRF subfigure, the top axis shows vector amplitudes of voxels in the left hemisphere and the bottom axis correspondingly in the right hemisphere. Each trace is from the TRF of a single voxel; non-significant time points are shown in gray, while significant time points are shown in red (sentence TRF) or blue (equation TRF). The duration plotted corresponds to that of a sentence or equation, plus 350 ms; because of the fixed presentation rate, the first 350 ms (shown in gray) are identical to the last 350 ms. The large peak

at the end (and beginning) of each TRF may either be ascribed to processing of the completion of the sentence/equation, or to the onset of the new sentence sentence/equation, or both. Word and symbol onset times are shown in red and blue dashed lines respectively; it can be seen that response contributions associated with them have been successfully regressed out. Volume source space distributions for several peaks in the TRF amplitudes are shown in the inlay plots, with black arrows denoting current directions (peaks automatically selected as local maxima of the TRFs). Although most of the TRF activity is dominated by neural currents in the left temporal lobe, the equation TRFs show more bilateral activation.

Differences in sentence and equation processing were more readily visible in the TRFs for the cocktail party conditions. The test for consistent vector direction revealed similar results to the single speaker conditions (sentence TRF $T^2_{max} = 5.15$, $p_{min} < 0.001$, equation TRF $T^2_{max} = 5.24$, $p_{min} < 0.001$) as shown in Fig. 4.6, however, the differences between sentences and equations were more pronounced, especially for the later peaks 410-600 ms after the onset of the last word or symbol. The peaks in the equation TRF were localized to left motor and parietal regions and right inferior frontal areas that are associated with arithmetic processing. This strengthens the hypothesis that these late peaks indicate lagged higher-level processing of the completed equation and not early auditory/preparatory processing of the subsequent equation. Although the cortical localization of sentence TRF peaks remain consistent in left temporal areas throughout most of the time course, the equation TRF peaks show several distinct cortical patterns, and may reflect distinct processes. The equation TRF showed strong activity in bilateral IPS, superior parietal and motor areas, while sentence TRFs consistently localized

predominantly to regions near left auditory cortex, even more so than in the single speaker case. Therefore, selective attention in the cocktail conditions seems to highlight differences between arithmetic and language processing, and possible explanations are discussed section 4.5.6.

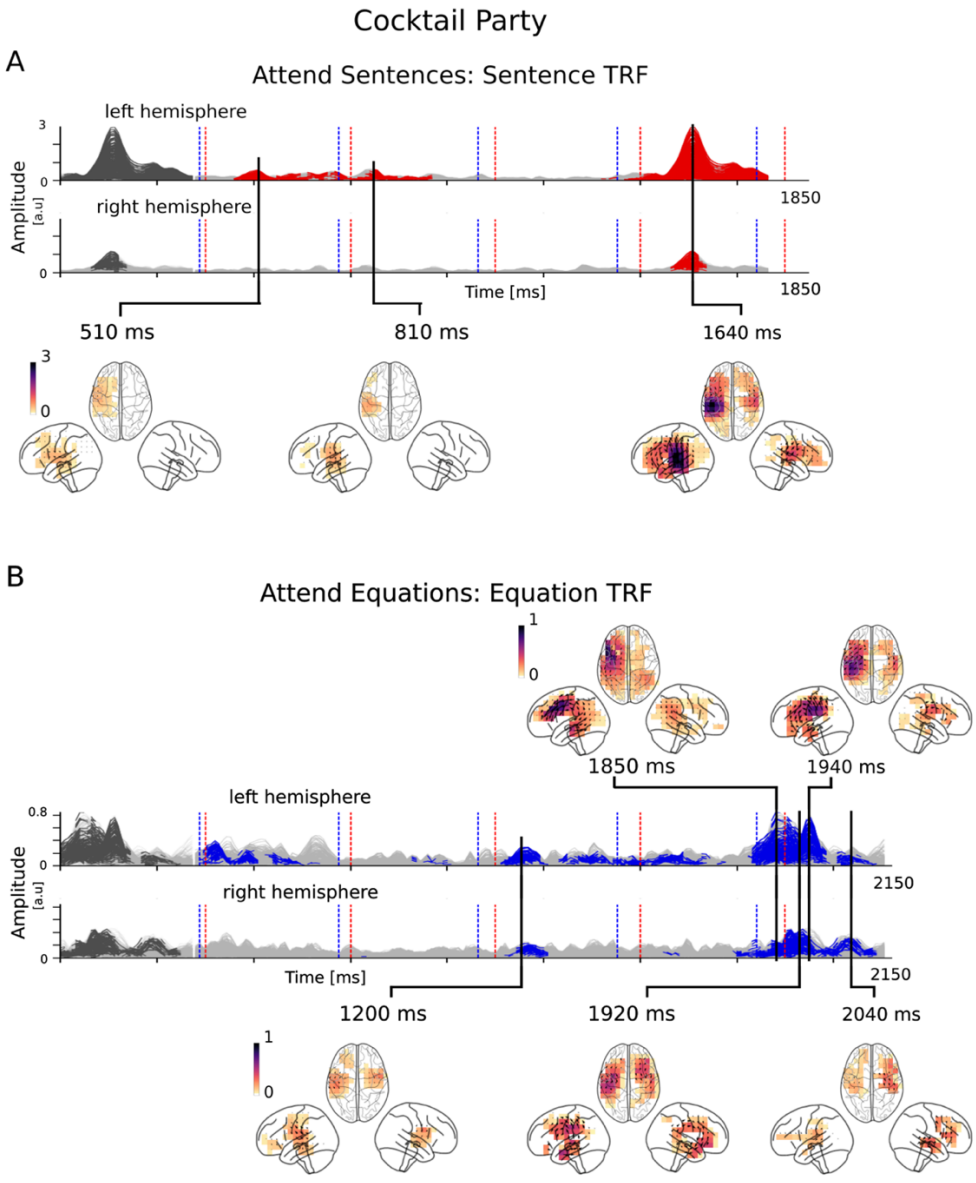

Figure 4.6. TRFs in the cocktail party conditions. Overlay plots of the TRF for each voxel averaged over subjects are shown as those in Fig. 4.5. Word and symbol onset times are shown in

red and blue dashed lines respectively and are marked in both sentence and equation TRFs since both stimuli were present in the cocktail party conditions; again, it can be seen that responses contributions associated with them have been successfully regressed out. Differences between sentence and equation TRFs arise at later time points, with sentence TRFs being predominantly near left temporal areas, while equation TRFs are in bilateral temporal, motor, and parietal regions.

4.4.5. Decoder analysis

To further help differentiate between the cortical processing of equations and sentences, two types of linear decoders were trained on neural responses. 1) Classifiers at each time point that learned weights based on the MEG sensor topography at that time point. 2) Classifiers at each voxel that learned weights based on the temporal dynamics of the response at that voxel. The former was used to contrast the dynamics of equation and sentence processing (Fig. 4.7A). For the single speaker conditions, all time points showed significant decoding ability across subjects (t_{max} = 11.3, p_{min} < 0.001), with higher prediction success (as measured by AUC) at longer latencies. For the cocktail party conditions, decoding ability was significantly above chance only at longer latencies (t_{max} = 6.45, p_{min} < 0.001). While subjects listened to the equations, the identity of the arithmetic operator (e.g., ‘plus’ vs. ‘times’ or ‘less’) was reliably decoded from the MEG sensor topography during the time points when the operator and the subsequent operand were presented (Fig. 4.7B). Note that decoding accuracy was significantly above chance for time points ~250-300 ms after the offset of the

operator. This is considerably late for decoding based on mere auditory responses to acoustic features of the operator.

Decoders at each voxel were also trained to differentiate attention to equations vs. sentences, based on the dynamics of the response at that voxel during the first or last words (Fig. 4.7C). The prediction success (as measured by *AUC*) was significant for large areas in the single speaker conditions both for first words ($t_{max} = 5.1$, $p_{min} < 0.001$) and last words ($t_{max} = 5.4$, $p_{min} < 0.001$) decoders. The *AUC* for first words decoders was significant for all regions in the left hemisphere except for areas in the inferior and middle temporal gyrus and all regions on the right hemisphere except the occipital lobe. The *AUC* for last words was significant for all regions in the left hemisphere and parts of frontal temporal lobes in the right hemisphere. For the cocktail party conditions the source-localized regions of significant prediction success were much more focal: the *AUC* was significant only in the IPS and superior parietal areas for both first words ($t_{max} = 5.3$, $p_{min} < 0.001$) and last words ($t_{max} = 4.3$, $p_{min} = 0.014$) decoders. These results suggest that the activity of voxels in left IPS and superior parietal areas is most useful for discriminating between attending to equations vs. sentences. Finally, decoders at each voxel were also trained to decode the attention condition (foreground vs. background) from the response to the entire sentence or equation (Fig. 4.7D). The *AUC* was significant in bilateral parietal areas for decoding whether arithmetic was in foreground vs. background ($t_{max} = 5.2$, $p_{min} < 0.001$), consistent with areas involved in arithmetic processing. For decoding whether language was in foreground vs. background, the *AUC* was significant ($t_{max} = 5.1$, $p_{min} = 0.002$) in left middle temporal

areas, consistent with higher level language processing, and bilateral superior parietal areas, consistent with attention networks that are involved in discriminating auditory stimuli. Therefore, the decoding analysis is able to detect different cortical areas that may be involved in attention to language and arithmetic.

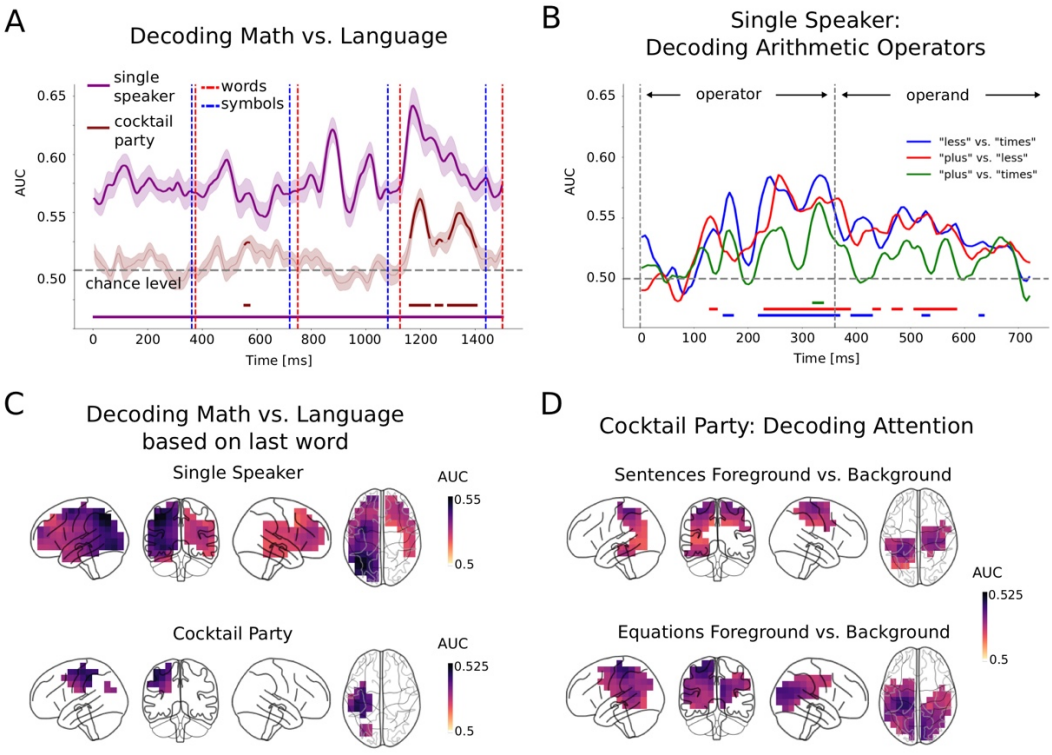

Figure 4.7. Decoding arithmetic and language processing. A. Performance of decoding attention condition (math vs. language) at each time point using MEG sensors for single speaker (purple) and cocktail party (brown). Prediction success is measured by *AUC*, which is plotted (mean and s.e.m. across subjects); time points where predictions are significantly above chance are marked by the horizontal bars at the bottom (every time point is significantly above chance for the single speaker case). The word and symbol onsets are also shown, and the decoding performance increases towards the end of the time window. **B.** Decoding arithmetic operators from sensor topographies. The time window of the operator and the subsequent operand was used for the 3 types of decoders. Time

intervals where predictions are significantly above chance are marked by the colored horizontal bars at the bottom: all 3 operator comparisons could be significantly decoded. **C.** Decoding math vs. language based on the last word. During the single speaker conditions, most of the brain is significant. However, for the cocktail party conditions, more focal significant decoding is seen in IPS and superior parietal areas. Decoding based on the first word resulted in similar results (not shown). **D.** Decoding attention in the cocktail party conditions (*AUC* masked by significance across subjects). The sentence responses in foreground and background were decoded in left middle temporal and bilateral superior parietal areas. The equation responses in foreground and background were decoded in bilateral parietal areas.

4.5. Discussion

We investigated the cortical locations and temporal dynamics of neural responses to spoken equations and sentences. Sentence responses consistently localized to left temporal areas. In contrast, equation responses consistently showed bilateral parietal activity, with other variations depending on analysis method (e.g., motor activity in TRFs). This may be due to different mechanisms involved in equation processing, although further investigation would be needed to support this claim. Cortical patterns consistent across different analysis methods (frequency domain, TRFs, decoders) are illustrated in schematic Fig. 4.8.

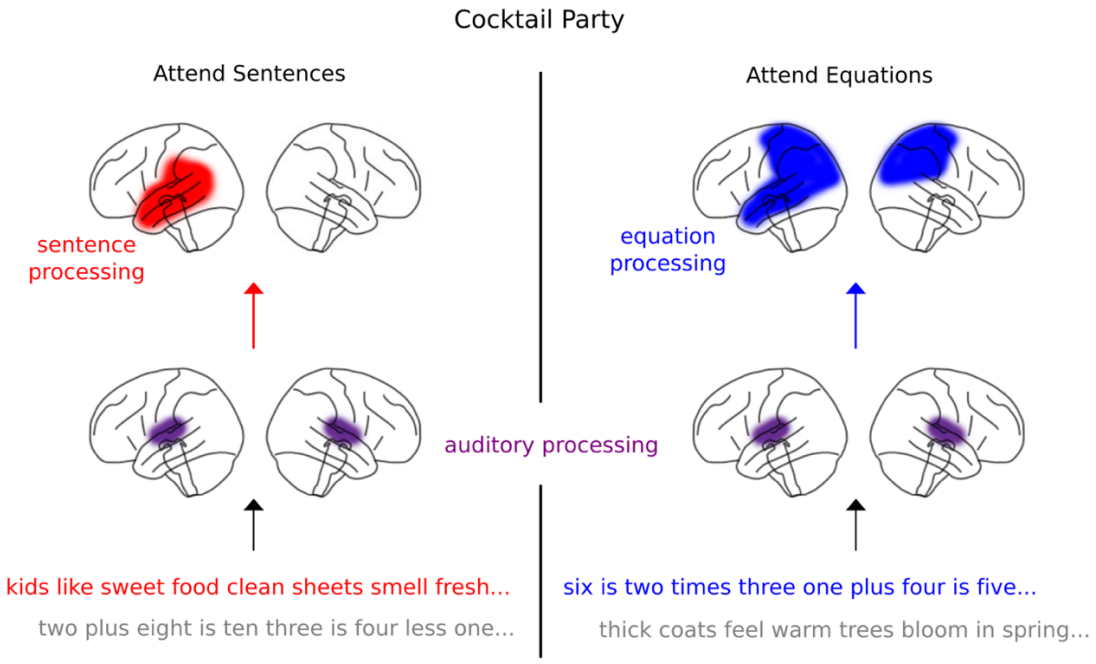

Figure 4.8. Schematic of cortical processing of sentences and equations. A schematic representation of sentence and equation processing is shown. Exemplars of both foreground and background of stimuli are shown at the bottom. The areas that were most consistent across all analysis methods (frequency domain, TRFs and decoders) are shown.

4.5.1. Sentence and equation rate responses

As expected, MEG responses to acoustic features source-localized to the bilateral auditory cortex, and sentence rate responses source-localized to the left temporal cortex, consistent with speech and language areas (Binder et al., 2009; Friederici, 2011, 2002; Hickok and Poeppel, 2007; Vandenberghe et al., 2002), similar to prior isochronous speech studies (Sheng et al., 2018). In contrast, equation rate responses localized to left parietal, temporal, and occipital areas. Arithmetic processing can

activate IPS and parietal (Dehaene et al., 2003), angular gyrus (Göbel et al., 2001), temporal (Tang et al., 2006), and even occipital areas (Harvey and Dumoulin, 2017; Maruyama et al., 2012), perhaps due to internal visualization (Zago et al., 2001). Equation responses also localized to the right temporal and parietal areas in cocktail party conditions, confirming that arithmetic processing is more bilateral than language processing (Amalric and Dehaene, 2019, 2018; Dehaene and Cohen, 1997). Critically, ANOVA analysis indicated that attended equation and sentence responses are significantly different. Unexpectedly, significant neural responses at the unattended sentence and equation rates were found in smaller temporal (consistent with language processing) and parietal (consistent with arithmetic processing) areas respectively. Some subjects may have been unable to sustain attention to the instructed stream for the entirety of this diotic stimulus and so briefly switched their attentional focus.

4.5.2. Left hemispheric dominance of equation responses

Equation responses were left dominant in both single speaker and cocktail party conditions. This could reflect left-lateralized language processing since equations were presented using speech. However, arithmetic processing may also show left dominance (Pinel and Dehaene, 2009), perhaps due to precise calculations (Dehaene, 1999; Pica et al., 2004) or arithmetic fact retrieval (Dehaene et al., 2003; Grabner et al., 2009). These fast-paced stimuli required rapid calculations, and may have resulted in increased reliance on rote memory, which activates left hemispheric areas (Campbell and Austin, 2002). Specific strategies employed for calculation may also result in left

lateralization—multiplication of small numbers is often performed using rote memory (Delazer et al., 1999; Fehr et al., 2007; Ischebeck et al., 2006), while subtraction is less commonly performed using memory and shows more bilateral activation (Prado et al., 2011; Schmithorst and Brown, 2004). Addition may recruit both these networks, depending on specific strategies utilized by individuals (Arsalidou and Taylor, 2011). We found no significant differences in equation responses when separated by operation type, perhaps because of individual variation in procedural calculation or retrieval strategies within the same operation (Tschentscher and Hauk, 2014). However, operation types were successfully decoded from the overall MEG signals (Fig. 4.8B), consistent with prior work (Pinheiro-Chagas et al., 2019), although not as robustly as decoding stimulus type or attention. Overall, left-hemispheric dominance of equation responses is supported by a combination of speech processing, precise calculations, and arithmetic fact retrieval.

4.5.3. Cortical correlates of behavioral performance

Neural responses to sentence, equation, word, and symbol rates were correlated with performance in detecting deviants, consistent with language-only isochronous studies (Ding et al., 2017). Sentence responses correlated with behavior in language areas, such as left auditory cortex, superior and middle temporal lobe and angular gyrus (Binder et al., 2009; Karuza et al., 2013; Price, 2000) in both single speaker and cocktail party conditions. In contrast, equation responses correlated with behavior in cocktail party conditions in posterior parietal areas, which are known to predict competence and

performance in numerical tasks (Grabner et al., 2007; Lasne et al., 2019; Lin et al., 2019, 2012). The lack of significant behavioral correlations for equation responses in single speaker conditions may be due to several subjects performing at ceiling; equations had a restricted set of only 14 unique symbols, and the presence of the 'is' symbol in every equation might be structurally useful in tracking equation boundaries. Unexpectedly, behavioral correlations were also found for background symbol rate responses in parietal and occipital areas (and for background word rate responses in a small parietal region). Some studies show that acoustic features of background speech may also be tracked (Brodbeck et al., 2020a; Fiedler et al., 2019). Since background word and symbol rates were present in the stimulus acoustics, increased effort or attention could enhance both behavioral performance and auditory responses at these rates. Representations of the background could enhance attentional selectivity in challenging cocktail party environments as suggested by Fiedler et al., 2019. However, note that sentence and equation responses were only correlated with behavior when attended. Overall, behavioral correlations in temporal and parietal regions suggest that these responses may reflect improved comprehension due to neural chunking of speech structures or successful calculations (Blanco-Elorrieta et al., 2019; Chen et al., 2020; Jin et al., 2020; Kaufeld et al., 2020; Teng et al., 2020).

4.5.4. Dynamics of arithmetic and language processing

TRFs for equation and sentence onsets were jointly estimated along with speech envelopes and word/equation onsets in order to regress out auditory responses,

analogous to prior work with linguistic and auditory TRFs (Brodbeck et al., 2018a, 2018b; Broderick et al., 2018). Isochronous speech studies have found slow rhythmic activity (Zhang and Ding, 2017), which did not appear in our TRFs, perhaps due to implicit high-pass filtering (boosting favors sparse TRFs). Instead, we found large TRF peaks at sentence/equation boundaries. Prior studies have found late evoked responses specific to numbers and equations (Avancini et al., 2015). Large peaks appeared in both sentence and equation TRFs 410–600 ms after the onset of the last word/symbol and may reflect processing of the completion of the sentence/equation. Sentence TRF peaks localized to left temporal areas, while equation TRF peaks showed activity in bilateral parietal and temporal areas involved in numerical processing (Abd Hamid et al., 2011; Amalric and Dehaene, 2018), and motor areas, perhaps reflecting procedural calculation strategies (Tschentscher and Hauk, 2014). The peak latencies were similar to prior arithmetic ERP studies (Iguchi and Hashimoto, 2000; Iijima and Nishitani, 2017). These sentence and equation TRF peaks may reflect several mechanisms; both shared (language processing, decision making), and separate (semantic vs. arithmetic processing), and further work is needed to disentangle these mechanisms. Finally, the cortical patterns of TRF peaks showed more differences in the cocktail party than the single speaker conditions, suggesting that selective attention focuses the underlying cortical networks.

4.5.5. Decoding equation and sentence processing

Numbers and arithmetic operations have been previously decoded from cortical responses (Eger et al., 2009; Pinheiro-Chagas et al., 2019). In this study, the attended stimulus type (sentences or equations) was reliably decoded in single speaker conditions in several cortical regions, perhaps due to highly correlated responses across cortex for this task. In contrast, decoding accuracy during cocktail party conditions was significant in left IPS and superior parietal areas, suggesting that these regions are most important for discriminating between arithmetic and language processing. Both the attend-equations and the attend-sentences states could be decoded from bilateral superior parietal areas, perhaps due to general attentional networks in fronto-parietal areas, or attentional segregation of foreground and background speech based on pitch or gender (Hill and Miller, 2010; Kristensen et al., 2013). Additionally, decoding the attend-equations state was significant in bilateral parietal areas, consistent with arithmetic processing, while decoding the attend-sentences state was significant in the left middle temporal lobe, consistent with language processing (Hickok and Poeppel, 2007). Overall, MEG responses contain enough information to decode arithmetic vs. language processing, selective attention, and arithmetic operations.

4.5.6. The cocktail party paradigm highlights distinct cortical processes

Differences in source-localized sentence and equation responses were more prominent in the cocktail party than in the single speaker conditions for all analyses

(frequency domain, TRFs and decoders). Responses to both stimuli presented simultaneously may have helped control for common auditory and pre-attentive responses. Abd Hamid et al., (2011) found fMRI activation in broader areas for spoken arithmetic with a noisy background than in quiet, perhaps due to increased effort. However, in our case, the background stimulus was not white noise, but rather meaningful non-mathematical speech. Our TRF analysis, which regresses out responses to background speech, as well as the selective attention task itself, may highlight specific cortical processes that best separate the arithmetic and language stimuli.

In summary, neural processing of spoken equations and sentences involves both overlapping and non-overlapping cortical networks. Behavioral correlations suggest that these neural responses may reflect improved comprehension and/or correct arithmetic calculations. Selective attention for equations focuses activity in temporal, parietal occipital and motor areas, and for sentences in temporal and superior parietal areas. This cocktail party paradigm is well suited to highlight the cortical networks underlying the processing of spoken arithmetic and language.

Chapter 5

A Comparison of Algorithms for Modelling Time-Locked Cortical Processing of Continuous Speech

5.1. Abstract

The Temporal Response Function (TRF) is a linear model of time-locked M/EEG activity to continuous stimuli that has proved to be successful in investigating cortical processing of continuous speech. TRFs to speech envelopes often have distinct components that are comparable to components found in Evoked Response Potentials (ERPs). Task and group differences in the amplitudes and latencies of these TRF components have provided several insights into speech processing. However, these component characteristics may depend on the specific TRF algorithm employed, and current methods often result in unreliable subject-specific components. In this work we provide a systematic comparison of TRF algorithms, in terms of their ability to estimate TRF components. Using both simulations and real MEG data, we compare two conventional algorithms, ridge regression and boosting, and two novel algorithms based on Orthogonal Matching Pursuit (OMP) and Expectation Maximization (EM). The proposed novel algorithms utilize prior knowledge of typical component characteristics to directly estimate component amplitudes and latencies. Comparisons were performed for single channel, multi-sensor, and source localized TRFs. The novel

algorithms outperformed the others in simulations, but did not perform well on real data, possibly because of unsuitable assumptions on component characteristics. Although ridge regression often resulted in the best model fit, all algorithms were comparable in terms of component estimation errors, especially at higher SNRs. Additional concerns such as sparsity and spurious TRF peaks are also discussed. This work highlights the importance of the choice of algorithm for estimating and detecting robust TRF components.

5.2. Introduction

The human brain time-locks to features of continuous speech, extracting meaningful information relevant to comprehension. Magnetoencephalography (MEG) and electroencephalography (EEG) are suitable methods to measure these time-locked responses, due to their high temporal resolution. Traditional methods for analyzing auditory responses involve averaging over several trials of repeated stimuli to estimate Evoked Response Potentials (ERPs) (Picton, 2013; Picton et al., 1974). However, exploring the complex mechanisms involved in speech processing requires non-repetitive, continuous speech stimuli of long duration, and averaging over many trials is no longer feasible. One method of analyzing responses to continuous stimuli uses linear models called Temporal Response Functions (TRFs), that seek to estimate the impulse response of the neural system to continuous stimuli (Ding and Simon, 2013, 2012a; Lalor and Foxe, 2010). These TRFs have response components that are similar to well-known auditory ERP components, and have been utilized to investigate

selective attention (Akram et al., 2016; Ding and Simon, 2012b; Miran et al., 2018), linguistic processing (Brodbeck et al., 2018a; Broderick et al., 2018) and age-related differences (Brodbeck et al., 2018b). However, though estimated TRFs display these canonical components at the group-average level, individual TRFs are much noisier and do not always have well-defined components. It is essential to detect robust response components on a per-subject level, in order to identify task effects, and to detect pathological cases.

In this work we compare TRF algorithms in terms of their ability to estimate TRF components, i.e., TRF peak latencies and amplitudes. Two of the most commonly used TRF estimation algorithms are ridge regression (Crosse et al., 2021, 2016) and boosting (David et al., 2007) , where the components of the TRF are greedily selected to decrease the mean square error (MSE) of the fit to the neural response. The former uses ℓ_2 regularization which leads to smooth TRFs with broad components, while the latter prioritizes sparsity in the TRF, leading to narrower, sharper components. However, it is not clear which of these methods is more accurate in estimating component latencies and amplitudes. Both ridge regression and boosting are agnostic to the morphology of neural responses. Given the fact that canonical auditory response components such as the M50 (~50 ms), M100 (~100-150 ms) and even M200 (~200-250 ms) are often present in TRFs to the speech envelope (Ding and Simon, 2012a), it is reasonable to incorporate this information during estimation.

Several methods have been proposed for estimating individual trial latencies and amplitudes for M/EEG evoked responses. The earliest ERP latency estimation methods

involved cross correlation with average response templates (Woody, 1967). More recent algorithms have utilized techniques such as Independent Component Analysis (Jung et al., 1999; Makeig et al., 2002), wavelet decomposition (Quiroga and Garcia, 2003), maximum likelihood estimation (de Munck et al., 2004; Jaskowski and Verleger, 1999), autoregressive models (Xu et al., 2009), Expectation Maximization (EM) (Limpiti et al., 2010) and Bayesian methods (Mohseni et al., 2010; Truccolo et al., 2003; Wu et al., 2014).

In this work, we also propose a novel TRF estimation algorithm that utilizes prior knowledge of the characteristics of neural responses, that is well suited to directly estimate component latencies, amplitudes and topographies. Given bounds on the latency ranges for each component, the proposed algorithm directly estimates component latencies and amplitudes using Orthogonal Matching Pursuit (OMP; Cai and Wang, 2011; Sieluzycki et al., 2009), and can be combined with the Expectation Maximization method (EM; Dempster et al., 1977; Limpiti et al., 2010) to directly estimate sensor topographies or source distributions for multidimensional TRFs.

A simulation study, as well as application of these algorithms to a real dataset, are reported and their performance is compared using single channel, sensor space and source localized TRFs. In addition to the conventional measure of model fit, the correlation between the actual and the predicted signal, several other performance metrics including errors in detecting peak amplitudes and latencies were used. Additionally, other considerations such as spurious TRF activity and missing components must also be taken into account when comparing these algorithms. This

work highlights the importance of estimating robust TRF components, discusses the strengths and weaknesses of widely used algorithms and proposes novel algorithms for TRF estimation that may provide robust and interpretable measures of time-locked response characteristics.

5.3. Methods

5.3.1. Ridge Regression

The TRF estimation problem is given by

$$y(t) = \sum_{k} \beta(k)x(t-k) + n(t) \tag{5.1}$$

Where $y(t)$ is the measured signal at one sensor for the t^{th} time point, x is the predictor variable, $\beta(k)$ is the TRF value for the k^{th} time lag and n is the noise. The above convolution equation describes the TRF as an impulse response of the neural system. This convolution can be reformulated as a regression:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{n} \tag{5.2}$$

Where $\mathbf{y}\epsilon\mathbb{R}^T$ is the vector of the measured signal for T time points, $\boldsymbol{\beta}\epsilon\mathbb{R}^K$ is the corresponding TRF over K time lags, $\mathbf{n}\epsilon\mathbb{R}^T$ is the noise and $\mathbf{X}\epsilon\mathbb{R}^{T\times K}$ is the Toeplitz matrix formed by lagged predictor values. The well-known ridge regression algorithm seeks to minimize the following cost function,

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_2^2 \tag{5.3}$$

Therefore, ridge regression seeks to minimize both the error between the actual and predicted signals and a regularization term on the TRF coefficients. The solution is given by,

$$\boldsymbol{\beta} = (\mathbf{X^T X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y} \tag{5.4}$$

The regularization parameter λ must be selected carefully. Here we use a nested cross-validation scheme to tune this parameter. Ridge regression results in smooth TRFs and can be used independently at multiple sensors to estimate TRFs for multi-channel data.

5.3.2. Boosting

The Boosting algorithm solves the TRF problem using a greedy coordinate descent. In brief, the algorithm starts from an all-zero TRF and incrementally adds small, fixed values to the TRF that lead to the largest decrease in the error measure at each iteration. In this work, we use the same error measure as in the regression problem (ℓ_2 norm, but without regularization), and stop the iterations when the Pearson correlation between the actual and predicted signals does not improve. TRFs at each sensor are computed independently and the complete algorithm is given below

Algorithm 5.1. Boosting

Inputs: Single sensor data $\mathbf{y}\epsilon\mathbb{R}^T$, and the predictor $\mathbf{x}\epsilon\mathbb{R}^T$

1: Split $\mathbf{y}$, $\mathbf{x}$ into training ($\mathbf{y}^{Tr}$, $\mathbf{x}^{Tr}$), validation ($\mathbf{y}^{V}$, $\mathbf{x}^{V}$) and testing subsets ($\mathbf{y}^{Te}$, $\mathbf{x}^{Te}$)

2: Select the step-size δ and convergence tolerance ε.

3: Initialize the all-zero TRF $\boldsymbol{\beta} = \mathbf{0}$ with K time lags. i.e., each element $\beta(k) = 0$

4: **repeat** for $i = 1, 2, \ldots$

5: For the lags $\tau \in \{1.., K\}$ and signs $\zeta\epsilon\{-1,1\}$ Define

$$\Delta\beta_{\tau,\zeta}(k) = \begin{cases} \zeta\delta & if\ k = \tau \\ 0 & otherwise \end{cases}$$

6: Find the incremental change to the TRF that best reduces the error in the training set

$$\tau^*,\ \zeta^* = \underset{\tau\in\{1..,K\},\ \zeta\epsilon\{-1,1\}}{\operatorname{argmin}} \sum_{t=1}^{T}\sum_{k=1}^{K}\left[y^{Tr}(t) - x^{Tr}(t-k)\left(\beta(k) + \Delta\beta_{\tau,\zeta}(k)\right)\right]^2$$

7: Add this change to the current TRF

$$\beta(k) = \beta(k) + \Delta\beta_{\tau^*,\zeta^*}(k)$$

8: Compute the Pearson correlation between the predicted and actual signals on the validation set

$$corr^{(i)} = correlation(\mathbf{y}^{V},\ \mathbf{x}^{V} * \boldsymbol{\beta})$$

9: **until** $corr^{(i)} - corr^{(i-1)} < \varepsilon$

Output: estimated TRF $\boldsymbol{\beta}$ and the correlation on the test dataset $correlation(\mathbf{y}^{Te},\ \mathbf{x}^{Te} * \boldsymbol{\beta})$

In practice, a dictionary of basis elements (e.g., Hamming windows) is used for the incremental additions to the TRF. Boosting estimates sparse TRFs and can be used independently at each sensor to estimate TRFs for multi-channel data.

5.3.3. Orthogonal Matching Pursuit (OMP)

The OMP algorithm searches for TRF components within predefined latency windows and directly estimates them. Assuming there are j=1 to J components (e.g., J=3 for M50, M100, M200 components), the TRF model is now given by a modified version of Eq. 5.1.

$$\mathbf{y} = \sum_{j} a_j \mathbf{X}\mathbf{c}_j + \mathbf{n} \tag{5.5}$$

Where $a_j \in \mathbb{R}$ and $\mathbf{c}_j \in \mathbb{R}^{K\times 1}$ are the amplitude and waveform for the jth component. The component waveforms $\mathbf{c}_j$ are selected according to the component latency τ_j from a basis dictionary (e.g., hamming windows) that span the TRF lags (i.e., $\mathbf{c}_j$ is column number τ_j of the basis dictionary matrix). The OMP algorithm directly estimates the amplitudes and latencies $\{a_j\}$ and $\{\tau_j\}$. The complete algorithm is given on the following page.

The OMP algorithm estimates very sparse TRFs composed of only the required number of components with predefined waveforms chosen from a basis dictionary. The OMP algorithm can also be applied independently at each sensor to estimate TRFs for multi-channel data.

Algorithm 5.2. OMP

Inputs: Single sensor data $\mathbf{y} \epsilon \mathbb{R}^T$, the predictor $\mathbf{x} \epsilon \mathbb{R}^T$, the number of components J along with their corresponding latency windows $\mathrm{W}_j = \{k_{j1}, k_{j2}\}$

1: Initialize the set of components to the empty set; $\mathcal{C} = \emptyset$. Initialize the set of available component windows to include all the latency windows; $\mathcal{W} = \bigcup_j \mathrm{W}_j$

2: Set the residual to the actual signal $\mathbf{r} = \mathbf{y}$

3: **repeat** for j from 1 to J

4: Find the best component latency, considering both positive and negative components.

$$\tau^*, \zeta^* = \underset{\tau \in \mathcal{W},\ \zeta \in \{-1,1\}}{\operatorname{argmax}} \mathrm{r}^T (\zeta \mathbf{X} \tilde{\mathbf{c}}_\tau)$$

Where $\tilde{\mathbf{c}}_\tau$ corresponds to the basis component with latency τ

5: Add the new component to the set of components $\mathcal{C} = \mathcal{C} \cup \{\zeta^* \tilde{\mathbf{c}}_{\tau^*}\}$

6: Remove the corresponding window from the set of available windows $\mathcal{W} = \mathcal{W} \setminus \widetilde{W}_{\tau^*}$

7: Re-estimate the amplitudes of all the components using the least squares method

$$\boldsymbol{a} = \underset{\boldsymbol{a} \in \mathbb{R}^j}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{X}_c \boldsymbol{a}\|$$

where $\mathbf{X}_c = [\mathbf{X}\mathbf{c}_1, \ldots, \mathbf{X}\mathbf{c}_j] \in \mathbb{R}^{N \times j}$

8: Calculate the new residual $\mathbf{r} = \mathbf{y} - \mathbf{X}_c \boldsymbol{a}$

Output: The TRF $\boldsymbol{\beta}$ given by the amplitudes $\boldsymbol{a} = [a_1, \ldots, a_J]$ and components in $\mathcal{C}$.

$$\boldsymbol{\beta} = \sum_{j=1}^{J} a_j \mathbf{c}_j$$

5.3.4. EM-OMP

The EM-OMP algorithm is an extension of the OMP algorithm for multidimensional TRFs. The goal is to directly estimate not only the amplitudes and latencies of TRF components, but also their sensor topographies using multi-channel data. This algorithm uses the Expectation Maximization (EM) method to iteratively estimate the component amplitudes and topographies in the E-step, and the latencies using OMP in the M-step. Given a predefined number of components and

corresponding latency windows, the EM-OMP multichannel TRF model is given by a modified version of Eq. 5.5.

$$\mathbf{Y} = \sum_j \mathbf{z}_j \left(\mathbf{X}\mathbf{c}_j\right)^{\mathrm{T}} + \mathbf{N} \tag{5.6}$$

Where $\mathbf{Y} \epsilon \mathbb{R}^{M \times T}$ is the measured data over M sensors and T time points, $\mathbf{z}_j \epsilon \mathbb{R}^{M}$ is the spatial topography of the j[th] component, $\mathbf{c}_j \epsilon \mathbb{R}^{K}$ is the temporal waveform of the j[th] component and $\mathbf{N} \epsilon \mathbb{R}^{M \times T}$ is the measurement noise. $\mathbf{X} \epsilon \mathbb{R}^{T \times K}$ is the predictor matrix with each column corresponding to lagged predictors. The component latency is given by τ_j and is related to Eq. 5.6 by the fact that $\boldsymbol{c}_j$ corresponds to column number τ_j in the TRF basis dictionary matrix. We assume the following priors,

$$\begin{gathered} \mathbf{z}_j \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{R}) \\ \mathbf{N} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{T \times T} \otimes \boldsymbol{\Lambda}) \end{gathered} \tag{5.7}$$

Where the temporal noise covariance is assumed to be the identity matrix and the spatial noise covariance is given by $\boldsymbol{\Lambda} \epsilon \mathbb{R}^{M \times M}$.

For the EM algorithm, we consider the spatial topographies $\mathcal{Z} = \{\mathbf{z}_j\}$ as the 'hidden' variables. The remaining parameters that need to be estimated are $\Theta = \{\tau_j, \boldsymbol{\mu}, \mathbf{R}, \boldsymbol{\Lambda}\}$. The data likelihood is given by

$$p(\mathbf{Y}|\mathcal{Z}; \Theta) \sim \mathcal{N}\left(\sum_j \mathbf{z}_j \mathbf{x}_j^T, \mathbf{I}_{T \times T} \otimes \boldsymbol{\Lambda}\right) \tag{5.8}$$

Where for convenience we have denoted $\mathbf{x}_j^T = \left(\mathbf{X}\mathbf{c}_j\right)^T$.

Assuming $\{\mathbf{z}_j\}$ are i.i.d., the complete data log likelihood is then given by

$$\mathcal{L}_c(\mathbf{Y},\mathcal{Z};\Theta) = \log p(\mathbf{Y}|\mathcal{Z};\Theta) + \sum_j \log p(\mathbf{z}_j)$$

$$\propto \frac{T}{2}\log|\mathbf{\Lambda}^{-1}| - \frac{1}{2}tr\left[\left(\mathbf{Y} - \sum_j \mathbf{z}_j\mathbf{x}_j^T\right)^T \mathbf{\Lambda}^{-1}\left(\mathbf{Y} - \sum_j \mathbf{z}_j\mathbf{x}_j^T\right)\right] \tag{5.9}$$

$$+\frac{J}{2}\log|\mathbf{R}^{-1}| - \frac{1}{2}\sum_j (\mathbf{z}_j - \boldsymbol{\mu})^T\mathbf{R}^{-1}(\mathbf{z}_j - \boldsymbol{\mu})$$

Hence, the Q-function is given by

$$Q(\Theta|\Theta^{(k)}) = \mathbf{E}_{\mathcal{Z}|\mathbf{Y};\Theta}[\mathcal{L}_c(\mathbf{Y},\mathcal{Z};\Theta)]$$

$$= \frac{T}{2}\log|\mathbf{\Lambda}^{-1}| - \frac{1}{2}tr[\mathbf{Y}^T\mathbf{\Lambda}^{-1}\mathbf{Y}] + tr\left[\mathbf{Y}^T\mathbf{\Lambda}^{-1}\left(\sum_j \mathbf{E}[\mathbf{z}_j]\mathbf{x}_j^T\right)\right]$$

$$-\frac{1}{2}\,tr\left[\left(\sum_j\sum_i \mathbf{x}_j^T\mathbf{x}_i\mathbf{E}[\mathbf{z}_j\mathbf{z}_i^T]\right)\mathbf{\Lambda}^{-1}\right] \tag{5.10}$$

$$+\frac{J}{2}\log|\mathbf{R}^{-1}|$$

$$-\frac{1}{2}\sum_j \left(tr(\mathbf{E}[\mathbf{z}_j\mathbf{z}_j^T]\mathbf{R}^{-1}) - 2\boldsymbol{\mu}^T\mathbf{R}^{-1}\mathbf{E}[\mathbf{z}_j] + \boldsymbol{\mu}^T\mathbf{R}^{-1}\boldsymbol{\mu}\right)$$

The expectation is over the posterior distribution $p(\mathcal{Z}|\,\widetilde{\mathbf{Y}};\Theta) \propto \mathcal{L}_c(\mathbf{Y},\mathcal{Z};\Theta)$. Since this is quadratic in $\mathbf{z}_j$, the posterior for each $\mathbf{z}_j$ is normal.

$$p(\mathbf{z}_j|\,\widetilde{\mathbf{Y}};\Theta) = \mathcal{N}(\bar{\mathbf{z}}_j, \mathbf{S}_j) \tag{5.11}$$

Using the properties of a Gaussian pdf, the mean $\bar{\mathbf{z}}_j$ of the Gaussian is given by setting the derivative to zero.

$$\boldsymbol{\Lambda}^{-1}\left(\mathbf{Y}-\sum_{i}\mathbf{z}_i\mathbf{x}_i^T\right)\mathbf{x}_j-\mathbf{R}^{-1}(\mathbf{z}_j-\boldsymbol{\mu})=\ 0 \tag{5.12}$$

The covariance $\mathbf{S}_j$ of the Gaussian is given by the inverse of the Hessian.

$$\mathbf{S}_j=\left(\mathbf{x}_j^T\mathbf{x}_j\boldsymbol{\Lambda}^{-1}+\mathbf{R}^{-1}\right)^{-1} \tag{5.13}$$

Using Eq. 5.12 and Eq. 5.13 the mean is given by

$$\bar{\mathbf{z}}_j=\mathbf{S}_j(\boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{Y}}\mathbf{x}_j+\mathbf{R}^{-1}\boldsymbol{\mu}) \tag{5.14}$$

$$where\quad \widetilde{\mathbf{Y}}=\mathbf{Y}-\sum_{i\neq j}\mathbf{z}_i\mathbf{x}_i^T \tag{5.15}$$

Note that the solution for each $\bar{\mathbf{z}}_j$ are coupled through $\widetilde{\mathbf{Y}}$ and a system of linear equations must be solved. However, in practice, solving each $\bar{\mathbf{z}}_j$ while keeping the others fixed leads to an approximate solution and avoids instabilities and computations of large matrix inversions. Therefore, we use Eq. 5.14 for each $\bar{\mathbf{z}}_j$ to solve for the posterior mean. The relevant terms in the Q-function can now be substituted as follows

$$\begin{gathered}\mathbf{E}[\mathbf{z}_j]=\bar{\mathbf{z}}_j\\ \mathbf{E}\left[\mathbf{z}_j\mathbf{z}_j^T\right]=\mathbf{S}_j+\bar{\mathbf{z}}_j\bar{\mathbf{z}}_j^T\\ \mathbf{E}\left[\mathbf{z}_j\mathbf{z}_i^T\right]=\bar{\mathbf{z}}_j\bar{\mathbf{z}}_i^T\ \ for\ \ i\neq j\end{gathered} \tag{5.16}$$

For the M-step, we maximize the Q-function w.r.t to the other parameters $\Theta=\{\tau_j,\boldsymbol{\mu},\mathbf{R},\boldsymbol{\Lambda}\}$. We use the Conditional Maximization method (Meng and Rubin, 1993) whereby we sequentially maximize over each one of these parameters while holding the others fixed at their previous values. Maximization updates are found by setting the partial derivatives of the Q-function to zero

For $\boldsymbol{\mu}$, the relevant terms of the Q-function are:

$$\begin{aligned} &-\frac{1}{2}\sum_j\left(-2\boldsymbol{\mu}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{E}\left[\mathbf{z}_j\right]+\boldsymbol{\mu}^T\mathbf{R}^{-1}\boldsymbol{\mu}\right) \\ &=-\frac{1}{2}\sum_j\left(-2\boldsymbol{\mu}^{\mathrm{T}}\mathbf{R}^{-1}\bar{\mathbf{z}}_j+\boldsymbol{\mu}^T\mathbf{R}^{-1}\boldsymbol{\mu}\right) \end{aligned} \tag{5.17}$$

Setting the derivative to zero, we find the update

$$\boldsymbol{\mu}=\frac{1}{J}\sum_j\bar{\mathbf{z}}_j \tag{5.18}$$

For $\mathbf{R}$, the relevant terms of the Q-function are:

$$\begin{aligned} \frac{J}{2}\log|\mathbf{R}^{-1}|-\frac{1}{2}\sum_j\Big(tr\big((\mathbf{S}_j+\bar{\mathbf{z}}_j\bar{\mathbf{z}}_j^T)\mathbf{R}^{-1}\big)-\boldsymbol{\mu}^{\mathrm{T}}\mathbf{R}^{-1}\bar{\mathbf{z}}_j-\bar{\mathbf{z}}_j^T\mathbf{R}^{-1}\boldsymbol{\mu} \\ +\boldsymbol{\mu}^T\mathbf{R}^{-1}\boldsymbol{\mu}\Big) \end{aligned} \tag{5.19}$$

Hence, setting the derivative w.r.t. $\mathbf{R}^{-1}$ to zero,

$$\mathbf{R}=\frac{1}{JM}\sum_j\left(\mathbf{S}_j+\bar{\mathbf{z}}_j\bar{\mathbf{z}}_j^T-\boldsymbol{\mu}\bar{\mathbf{z}}_j^T-\bar{\mathbf{z}}_j\boldsymbol{\mu}^T+\boldsymbol{\mu}\boldsymbol{\mu}^T\right) \tag{5.20}$$

For $\boldsymbol{\Lambda}$**,** the relevant terms of the Q-function are:

$$\begin{aligned} &\frac{T}{2}\log|\boldsymbol{\Lambda}^{-1}|-\frac{1}{2}tr[\mathbf{Y}^T\boldsymbol{\Lambda}^{-1}\mathbf{Y}]+\frac{1}{2}tr\left[\mathbf{Y}^T\boldsymbol{\Lambda}^{-1}\left(\sum_j\bar{\mathbf{z}}_j\mathbf{x}_j^T\right)\right] \\ &+\frac{1}{2}tr\left[\left(\sum_j\bar{\mathbf{z}}_j\mathbf{x}_j^T\right)^T\boldsymbol{\Lambda}^{-1}\mathbf{Y}\right] \\ &-\frac{1}{2}\sum_j\left(\mathbf{x}_j^T\mathbf{x}_j tr\left[\left(\mathbf{S}_j+\bar{\mathbf{z}}_j\bar{\mathbf{z}}_j^T\right)\boldsymbol{\Lambda}^{-1}\right]+\sum_{i\neq j}\mathbf{x}_j^T\mathbf{x}_i tr\left[\bar{\mathbf{z}}_j\bar{\mathbf{z}}_i^T\boldsymbol{\Lambda}^{-1}\right]\right) \end{aligned} \tag{5.21}$$

Calculating the derivative w.r.t. $\mathbf{\Lambda}^{-1}$ using the same methods and equating to zero:

$$\mathbf{\Lambda} = \frac{1}{T}\left[\mathbf{Y}\mathbf{Y}^T - \mathbf{Y}\left(\sum_j \bar{\mathbf{z}}_j \mathbf{x}_j^T\right)^T - \left(\sum_j \bar{\mathbf{z}}_j \mathbf{x}_j^T\right)\mathbf{Y}^T + \sum_j \left(\mathbf{x}_j^T \mathbf{x}_j \left(\mathbf{S}_j + \bar{\mathbf{z}}_j \bar{\mathbf{z}}_j^T\right)^T + \sum_{i \neq j} \mathbf{x}_j^T \mathbf{x}_i \bar{\mathbf{z}}_i \bar{\mathbf{z}}_j^T\right)\right] \tag{5.22}$$

For τ_j, the relevant terms involving $\mathbf{x}_j$ in the Q-function are

$$tr\left[\widetilde{\mathbf{Y}}^T \mathbf{\Lambda}^{-1} \bar{\mathbf{z}}_j \mathbf{x}_j^T\right] - \frac{1}{2}\left[\mathbf{x}_j^T \mathbf{x}_j tr\left[\left(\mathbf{S}_j + \bar{\mathbf{z}}_j \bar{\mathbf{z}}_j^T\right)\mathbf{\Lambda}^{-1}\right] + \sum_{i \neq j} \mathbf{x}_j^T \mathbf{x}_i tr\left[\bar{\mathbf{z}}_j \bar{\mathbf{z}}_i^T \mathbf{\Lambda}^{-1}\right]\right] \tag{5.23}$$

We assume that the second term doesn't depend on τ_j (i.e., $\mathbf{x}_j^T \mathbf{x}_j$ and $\mathbf{x}_j^T \mathbf{x}_i$ are similar for all rows of the lagged stimulus matrix since they have similar vector norms). Therefore, the only term that depends on the latency is the first term $tr\left[\widetilde{\mathbf{Y}}^T \mathbf{\Lambda}^{-1} \bar{\mathbf{z}}_j \mathbf{x}_j^T\right]$, which is the correlation between the whitened measurements and predictions. To maximize this term, we use the OMP algorithm, with appropriate modifications to include spatial topographies. The complete algorithm is provided below.

Algorithm 5.3. EM-OMP

Inputs: Multichannel data $\mathbf{Y} \in \mathbb{R}^{M \times T}$, $\mathbf{X} \in \mathbb{R}^{T \times K}$ the number of components J along with their corresponding latency windows $\mathrm{W}_j = \{k_{j1}, k_{j2}\}$

1: Initialize the parameters $\Theta^0 = \{\tau_j^0, \boldsymbol{\mu}^0, \mathbf{R}^0, \boldsymbol{\Lambda}^0\}$.

2: **repeat** for t from 1 to convergence

3: E-step: Estimate the spatial topographies $\bar{\mathbf{z}}_j$ using Eq. 5.14

4: CM-steps: Estimate parameters $\boldsymbol{\mu}^t, \mathbf{R}^t, \boldsymbol{\Lambda}^t$ using Eq. 5.18 - 5.22

5: CM-step: Estimate the latencies τ_j^t using OMP as given below

6: Initialize the components to the empty set $\mathcal{C} = \phi$

7: Initialize the residual $\widetilde{\mathbf{Y}} = \mathbf{Y}$

8: **repeat** for j = 1 until J

9: Find the best component latency that maximizes the following

(a search over integer latencies is used)

$$\tau_j = \underset{k \in (k_{j1}, k_{j2})}{\operatorname{argmax}} \; tr\left(\widetilde{\mathbf{Y}}^T \boldsymbol{\Lambda}^{-1} \bar{\mathbf{z}}_j (\mathbf{X}\mathbf{g}_k)^T\right)$$

10: Update $\mathcal{C} = \mathcal{C} \cup \{\mathbf{c}_j\}$ with $\mathbf{c}_j = \mathbf{g}_{\tau_j}$

11: Update the amplitudes of each component

$$\boldsymbol{a} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{y}$$

where $\mathbf{y} \in \mathbb{R}^{NM \times 1}$ is the vectorized whitened data $\mathbf{y} = vec\left(\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{Y}\right)$

and $\mathbf{D} \in \mathbb{R}^{NM \times j}$ has columns $\boldsymbol{d}_i = vec(\boldsymbol{\Lambda}^{-\frac{1}{2}}\bar{\mathbf{z}}_i x_i^T)$.

12: Update the topographies with the entries a_j of $\boldsymbol{a} \in \mathbb{R}^{j \times 1}$

$$\bar{\mathbf{z}}_j = a_j \bar{\mathbf{z}}_j$$

13: Update the residual $\widetilde{\mathbf{Y}} = \mathbf{Y} - \sum_{i=1}^{j} \bar{\mathbf{z}}_i \mathbf{x}_i^T$

Output: The estimated TRF $\boldsymbol{\beta}$ given by the estimated component topographies $\bar{\mathbf{z}}_j$ and latencies τ_j. The component amplitudes are given by $a_j = \max\left(abs\left(\bar{\mathbf{z}}_j\right)\right)$. The TRF is given by $\boldsymbol{\beta} = \sum_{j=1}^{J} \mathbf{z}_j \mathbf{c}_j^T$

5.3.5. Algorithm Implementation

The ridge regression, OMP and EM-OMP algorithms were implemented in python using scipy (Virtanen et al., 2020), and eelbrain (Brodbeck et al., 2021a) was used for boosting. A Hamming window basis of width 50 ms was used for the boosting, OMP and EM-OMP algorithms. The same nested 4-fold cross validation procedure was followed for all algorithms to allow for a fair comparison. The data was split into 4 splits, with 1 split for testing, 1 for validation and 2 for training. The validation and training splits were permuted for each test split in a nested fashion. The boosting TRF was fit on the training data, and the validation data was used to check for convergence and terminate the algorithm. The training data was used to fit the ridge regression TRF over several regularization parameters and the parameter that gave the highest correlation on the validation data was selected. The OMP TRF was fit on the training data, and the EM-OMP TRF was fit on the training data with the validation data being used to check for convergence and terminate the iterations. The predicted test signal was computed by convolving the average TRF over all training splits with the appropriate test predictor and the model fit was calculated as the Pearson correlation between this predicted signal and the actual test signal.

The latency windows for the OMP and EM-OMP algorithms were fixed to be 30-80 ms, 90-190 ms and 200-300 ms, to roughly correspond to typical latencies of the M50, M100 and M200 components. The EM-OMP algorithm is sensitive to initializations and hence was initialized using the extracted components from the OMP algorithm applied at each sensor/source independently. To avoid instability and

convergence issues, the component covariance matrix **R** was assumed to be the identity matrix. The EM-OMP iterations were terminated if the correlation between the actual and the predicted signals of the validation dataset did not improve.

All of these algorithms can also be used to simultaneously fit TRFs to multiple predictors (e.g., foreground and background envelopes, or, envelopes and phoneme onset predictors). This can be done by concatenating the P predictor matrices $X_p \in \mathbb{R}^{T\times K}$ along the columns resulting in a new predictor matrix $X \in \mathbb{R}^{T\times KP}$. The resulting concatenated TRF can then be separated back into the TRFs specific to each predictor. In this work, we jointly fit TRFs to two predictors (speech envelopes) using a concatenated predictor matrix.

5.3.6. Simulation study

MEG responses were simulated in order to compare the performance of each algorithm. 3 minutes of MEG responses were simulated at 100 Hz sampling frequency for 30 pseudo-subjects as representative of a typical MEG experimental condition for TRF analysis. 3 types of simulations were done: single channel, sensor space (157 channels), and source space (245 sources in temporal lobe).

The single channel simulation consisted of data generated using the linear TRF model given in Eq. 5.1. The 1-10 Hz band-passed envelopes of two speech stimuli were used as the predictors, in order to simulate a cocktail party paradigm with both foreground and background speakers. These predictors were convolved with simulated

TRFs to form a one dimensional simulated MEG time series, comparable to a single M/EEG channel or the auditory component after an appropriate spatial filter. For each subject, the true TRF used for simulation was formed by placing hamming windows of 50 ms width at latencies in the range of 30-80 ms, 90-190 ms and 200-300 ms to roughly correspond to the M50, M100 and M200 components. The M100 component was given a negative sign, and the components were scaled and shifted according to randomized subject specific amplitudes and latencies. These randomized subject specific amplitudes and latencies were considered as the ground truth to evaluate the performance of each algorithm. Noise was added using the auditory component of real MEG data collected from 30 subjects listening to speech in quiet (previously published in Presacco et al., 2016a, 2016b). The first Denoising Source Separation (DSS) component was used for this auditory component (see Cheveigné and Simon, 2008 for details on DSS). The phase of this component was randomized, creating a noise signal that preserved the spectral properties of MEG neural responses, resulting in a realistic noise model. This noise signal was added to the simulated response at SNRs of -15, -20, -25 and -30 dB.

The sensor space simulation followed the same method to simulate TRFs, but in addition also used ground truth sensor topographies for each component. These topographies were constructed to be similar to typical auditory component sensor topographies. Gaussian noise was added to the topographies of each subject to simulate individual variability. This multichannel TRF was convolved with the envelope predictors to produce the measured multichannel signal according to Eq. 5.6. Noise was

added on a per channel basis using the method described above at SNRs of -20, -25, -30 and -35 dB (lower SNRs were used because unprocessed multichannel data is typically noisier than the extracted auditory component).

The envelope predictors used for these simulations were comprised of three repetitions of a one minute speech segment, similar to typical MEG recordings that use multiple trials in order to allow for extraction of auditory response components using DSS. Since these DSS components typically provide more meaningful TRFs than whole head sensor TRFs, the DSS algorithm was also applied to the simulated data and corresponding TRFs were calculated for the first 6 DSS components. These DSS TRFs were projected back into sensor space for subsequent analysis and for computing performance metrics.

The source space simulation was constructed using dipoles for each TRF component, under the assumption that these components arise from different sources in the cortex. The Freesurfer ico-4 surface source space of the 'fsaverage' brain was used (Fischl, 2012). An ROI in temporal lobe that included auditory cortex was used for this simulation ('aparc' labels 'transversetemporal' and 'superiortemporal'). The three TRF components were simulated using dipoles in Heschl's gyrus, Planum Temporale and Superior Temporal Gyrus in both hemispheres. These dipoles were projected onto the sensors using forward models from real data and back projected back onto source space with Minimum Norm Estimation (Hämäläinen and Ilmoniemi, 1994) using the mne-python software (A. Gramfort et al., 2014) to simulate the source localization procedure. The same procedure as the multichannel simulation was

followed for the rest of the simulation, with these back-projected component source distributions being used instead of the sensor topographies. MEG phase scrambled noise was added at SNRs of -15, -20, -25 and -30 dB in the same manner as described above for the sensor space simulations.

5.3.7. Experimental dataset

MEG data collected in a prior study (Presacco et al., 2016a, 2016b) was used for evaluating the performance of the algorithms on real data. The dataset consisted of MEG data collected from 40 subjects while they listened to speech from the narration of an audiobook. Subjects listened to two speakers simultaneously in a cocktail party experiment, but were asked to attend to only one speaker. The data was from the +3dB condition, where the foreground speaker was 3 dB louder than the background speaker. TRFs were estimated for two predictors: the foreground and background envelopes. Whole head sensor space TRFs (157 sensors) were computed for each algorithm on three minutes of data. Additionally, the DSS method was used on this data, which consisted of three repetitions of one minute speech segments. TRFs were computed for the first 6 DSS components. Finally, the MEG responses of this dataset were source localized using MNE onto the same surface source space ROI as that used for the simulations and source localized TRFs were also computed.

5.3.8. Performance metrics

For the simulation study, the performance of each algorithm was compared using several metrics; 1) Pearson correlation between the actual and predicted response (using the test data from cross-validation as discussed in section 5.3.5), 2) Pearson correlation between the estimated and ground truth TRF, 3) Absolute error of individual component latency estimates 4) Absolute error of individual component amplitude estimates, 5) Spurious TRF activity given by % power in the estimated TRF after 300 ms (note that there is no activity in the ground truth TRF after 300 ms), 6) Number of missing components 7) Sensor topography error given by the angle between the estimated and ground truth component topographies 8) Source distribution error given by the fractional mean squared error between the estimated and ground truth component source distributions. To calculate the above metrics, the TRF components were found using automatic peak selection in the appropriate latency windows (M50: 30-80 ms, M100: 90-190 ms, M200: 200-300 ms).

5.4. Results

5.4.1. Simulation: single channel TRFs

Single channel TRFs were simulated and the ridge regression, boosting and OMP algorithms were compared in terms of several performance metrics (see Table 5.1, Fig. 5.1). The OMP algorithm performed the best in most measures, while ridge regression and boosting performed comparably. The boosting algorithm failed to detect more

components compared to ridge regression, while the latter had more spurious activity after 300 ms (when there was no activity in the ground truth TRF). This pattern is highlighted in the individual TRFs for a representative subject shown in Fig. 5.1A.

Table 5.1. Performance comparison for single channel simulations

SNR	Algorithm	Correlation (Pearson r)	TRF correlation (Pearson r)	Latency error (m.s.)	Amplitude error (a.u.)	Spurious activity (% power > 300ms)	% Missing components
-15	Boosting	0.26 [0.05]	0.79 [0.05]	13.4 [3]	0.71 [0.33]	**0.7 [1]**	9.5 [4.9]
	Ridge	0.27 [0.05]	**0.82 [0.04]**	11.3 [3]	0.39 [0.18]	6.1 [1.7]	**0.86 [1]**
	OMP	**0.28 [0.05]**	0.77 [0.04]	**10.6 [1.8]**	**0.34 [0.15]**	--	--
-20	Boosting	0.14 [0.02]	0.69 [0.07]	17.6 [3]	0.93 [0.37]	**3.2 [2]**	17.5 [5.6]
	Ridge	0.15 [0.02]	**0.72 [0.05]**	15.3 [3]	0.65 [0.27]	12.3 [2.4]	**3.2 [1.8]**
	OMP	**0.16 [0.02]**	0.71 [0.05]	**13 [1.9]**	**0.49 [0.2]**	--	--
-25	Boosting	0.07 [0.02]	0.53 [0.09]	22.2 [1.8]	1.04 [0.35]	**12.9 [5.9]**	26.4 [3.9]
	Ridge	0.089 [0.02]	0.56 [0.08]	20 [1.8]	1.19 [0.45]	22.2 [4.4]	**5.8 [2.4]**
	OMP	**0.09 [0.02]**	**0.62]0.05]**	**15.5 [1.4]**	**0.66 [0.26]**	--	--
-30	Boosting	0.04 [0.01]	0.34 [0.08]	26 [1.6]	1.07 [0.3]	**24.5 [5.7]**	32.8 [3.2]
	Ridge	**0.052 [0.01]**	0.36 [0.09]	24.8 [2.3]	1.39 [0.52]	31.3 [3.7]	**8.4 [2.7]**
	OMP	0.05 [0.01]	**0.5 [0.05]**	**18 [1.5]**	**0.83 [0.3]**	--	--

The mean [SD] over subjects is shown. The algorithm with the best performance for each metric at each SNR is highlighted in bold. OMP has neither spurious activity nor missing components by design since it estimates all the components using latency windows before 300 ms.

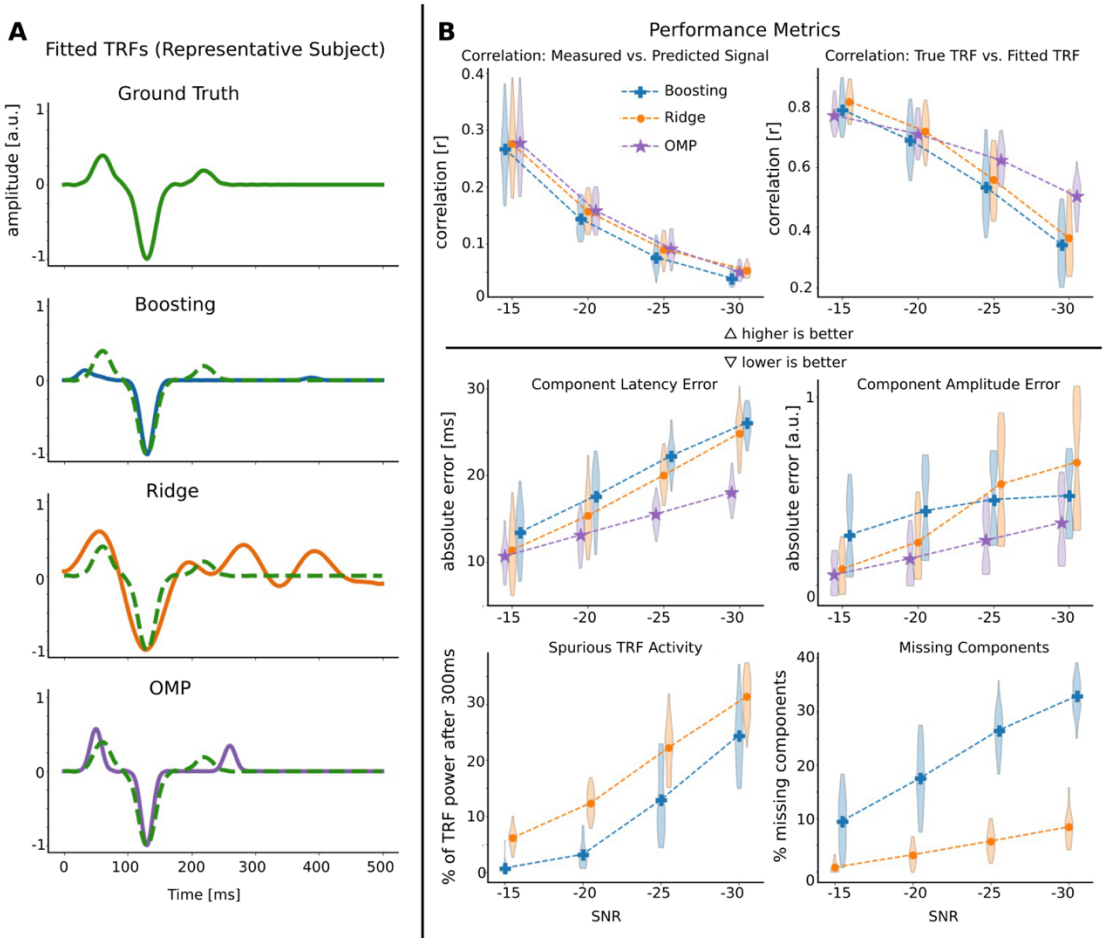

Figure 5.1. Performance comparison for single channel simulations. A. The fitted TRFs for a representative subject are shown. The ground truth TRF is shown as a dotted green line over the estimated TRFs. Boosting seems to miss some components, while ridge regression has more spurious activity. **B.** Algorithm comparison using the performance metrics. Violin plots over subjects are shown, with the symbols indicating the mean. Within each SNR condition, the algorithms are plotted in ascending order of their means from left to right. OMP does not have spurious activity or missing peaks by design and is not shown for the bottom two subplots. Ridge regression and boosting are comparable for most measures, while OMP seems to outperform the conventional algorithms in higher SNR cases.

5.4.2. Simulation: sensor space TRFs

Sensor space TRFs were simulated using realistic sensor topographies for TRF components, and the performance of each algorithm was compared (see Table 5.2, Fig. 5.2). TRFs were estimated independently at each sensor for the boosting, ridge regression and OMP algorithms, while the EM-OMP algorithm directly estimated component topographies. The EM-OMP algorithm performed the best in most measures, while ridge regression and boosting performed comparably. The sensor topographies estimated by boosting and OMP are much worse than those estimated by ridge regression and EM-OMP, which is to be expected given that the former are sparse algorithms fit at each sensor independently.

Table 5.2. Performance comparison for sensor space simulations

SNR	Algorithm	Correlation (Pearson r)	TRF correlation (Pearson r)	Latency error (m.s.)	Amplitude error (a.u.)	Spurious activity (% power > 300ms)	Topography error (rad)
-20	Boosting	0.069 [0.008]	0.603 [0.03]	7.3 [3.7]	0.229 [0.11]	**9.42 [1.9]**	0.86 [0.04]
	Ridge	0.077 [0.008]	0.595 [0.02]	9.7 [5.8]	0.105 [0.05]	18.66 [2.0]	0.62 [0.09]
	OMP	0.072 [0.007]	0.45 [0.03]	16.1 [6.7]	0.177 [0.07]	--	1.02 [0.09]
	EM-OMP	**0.085 [0.008]**	**0.82 [0.03]**	**2.8 [1.8]**	**0.05 [0.03]**	--	**0.54 [0.05]**
-25	Boosting	0.028 [0.004]	0.42 [0.02]	13.9 [8.41]	0.164 [0.07]	**19.39 [2.6]**	1.10 [0.07]
	Ridge	0.035 [0.004]	0.45 [0.02]	14.9 [6.44]	0.220 [0.1]	27.47 [1.9]	0.87 [0.09]
	OMP	0.032 [0.004]	0.35 [0.03]	16.5 [4.7]	0.179 [0.06]	--	1.13 [0.08]
	EM-OMP	**0.041 [0.004]**	**0.70 [0.04]**	**3.5 [1.9]**	**0.09 [0.03]**	--	**0.79 [0.08]**
-30	Boosting	0.007 [0.003]	0.25 [0.03]	21.0 [6.92]	0.262 [0.12]	**28.77 [2.7]**	1.27 [0.06]
	Ridge	0.013 [0.004]	0.29 [0.03]	21.9 [7.1]	0.282 [0.13]	33.76 [1.8]	1.12 [0.09]
	OMP	0.011 [0.004]	0.24 [0.02]	16.2 [6.3]	0.181 [0.05]	--	1.28 [0.08]
	EM-OMP	**0.015 [0.004]**	**0.50 [0.05]**	**7.1 [2.6]**	**0.16 [0.07]**	--	**1.02 [0.08]**
-35	Boosting	0.0002 [0.003]	0.147 [0.02]	24.7 [9.05]	0.217 [0.08]	**35.44 [2.7]**	1.38 [0.04]
	Ridge	**0.0044 [0.003]**	0.161 [0.05]	21.7 [7.7]	0.373 [0.18]	38.42 [2.1]	1.33 [0.1]
	OMP	0.0019 [0.003]	0.154 [0.02]	18.9 [6.1]	0.186 [0.06]	--	1.40 [0.05]
	EM-OMP	0.0039 [0.003]	**0.313 [0.04]**	**10.6 [2.8]**	**0.15 [0.05]**	--	**1.24 [0.05]**

The mean [SD] over subjects is shown. The algorithm with the best performance for each metric at each SNR is highlighted in bold. OMP and EM-OMP have no spurious activity by design since they estimate all the components using latency windows before 300 ms.

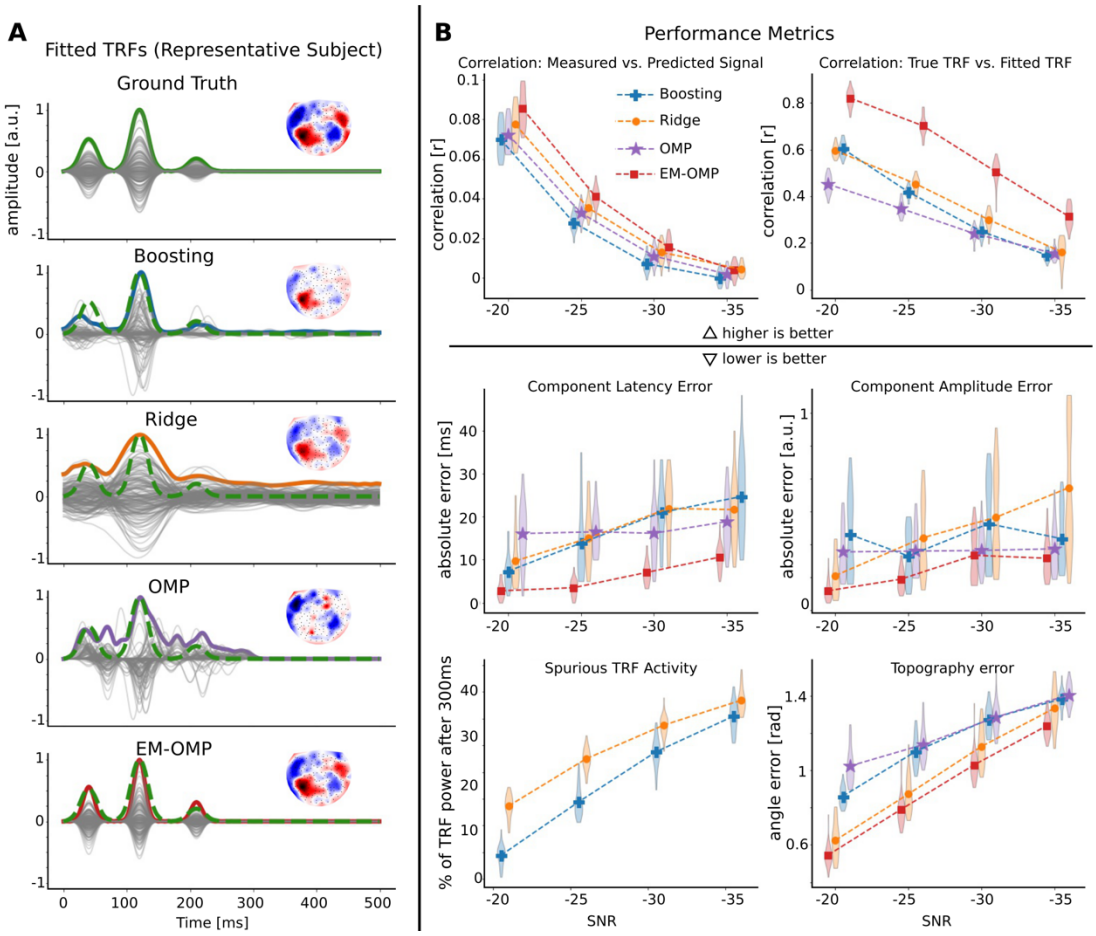

Figure 5.2. Performance comparison for sensor space simulations. A. The fitted TRFs for a representative subject are shown. The TRF at each sensor is plotted in gray, while the ℓ_2-norm over sensors is plotted as a colored thick line. The ℓ_2-norm of the ground truth TRF is shown as a dotted green line over the estimated TRFs. The sensor topography at the largest peak around 120 ms is shown as an inset. Although all methods find similar components, the sensor topographies for Boosting and OMP are much worse than those for ridge regression and EM-OMP, since the former are sparse algorithms. **B.** Algorithm comparison using the performance metrics, similar to those shown in the previous figure. Since there is no activity after 300 ms in EM-OMP and OMP TRFs by design, they are not plotted in the spurious activity subplot. EM-OMP outperforms the others in most measures.

5.4.3. Simulation: DSS TRFs

The DSS algorithm was applied to the simulated sensor space TRFs, in order to extract spatial filters corresponding to auditory response components. The algorithms were fit on the first 6 DSS components, and the resulting TRFs were projected back onto the sensor space and compared using the same performance metrics (see Table 5.3, Fig. 5.3). Performance increased greatly in all cases, with ridge regression, boosting and EM-OMP having comparable results. Interestingly, EM-OMP did not have a significant advantage over the other algorithms, indicating that the conventional algorithms are just as suitable for low dimensional, denoised data.

Table 5.3. Performance comparison for DSS simulations

SNR	Algorithm	Correlation (Pearson r)	TRF correlation (Pearson r)	Latency error (m.s.)	Amplitude error (a.u.)	Spurious activity (% power > 300ms)	Topography error (rad)
-20	Boosting	0.616 [0.05]	0.726 [0.05]	4.8 [2.4]	0.13 [0.05]	**0.07 [0.08]**	0.75 [0.1]
	Ridge	**0.619 [0.05]**	0.69 [0.03]	4.9 [4.2]	0.09 [0.05]	4.35 [1.1]	0.71 [0.1]
	OMP	0.56 [0.05]	0.39 [0.14]	17.2 [11.4]	0.2 [0.07]	--	0.96 [0.2]
	EM-OMP	0.618 [0.05]	**0.736 [0.06]**	**2.6 [2]**	**0.08 [0.04]**	--	**0.73 [0.1]**
-25	Boosting	0.416 [0.06]	0.47 [0.08]	7.3 [6]	0.2 [0.08]	**0.33 [0.35]**	1.09 [0.1]
	Ridge	**0.42 [0.06]**	0.43 [0.06]	7.6 [5.8]	**0.16 [0.08]**	11.39 [3.1]	1.06 [0.1]
	OMP	0.37 [0.05]	0.25 [0.09]	16.8 [9.8]	0.22 [0.08]	--	1.23 [0.1]
	EM-OMP	0.418 [0.06]	**0.48 [0.09]**	**4.7 [3]**	0.17 [0.07]	--	**1.05 [0.1]**
-30	Boosting	0.2 [0.05]	0.166 [0.08]	12.9 [6.8]	0.3 [0.14]	**2.45 [1.5]**	1.436 [0.1]
	Ridge	**0.21 [0.05]**	0.15 [0.07]	12.7 [6.1]	0.23 [0.11]	19.96 [4.6]	**1.41 [0.1]**
	OMP	0.19 [0.05]	0.09]0.06]	17.4 [8.9]	**0.21 [0.08]**	--	1.47 [0.1]
	EM-OMP	0.2 [0.05]	**0.168 [0.08]**	**11.3 [5.7]**	0.23 [0.1]	--	1.432 [0.1]
-35	Boosting	0.07 [0.03]	0.036 [0.04]	18 [9.1]	0.37 [0.17]	**14.09 [7.0]**	1.54 [0.1]
	Ridge	**0.08 [0.02]**	0.028 [0.09]	22.7 [8.3]	0.27 [0.12]	33.60 [5.7]	**1.526 [0.08]**
	OMP	0.06 [0.02]	0.037 [0.04]	18.6 [6.2]	**0.23 [0.3]**	--	1.53 [0.1]
	EM-OMP	0.07 [0.03]	**0.039 [0.04]**	**15.4 [5.4]**	0.24 [0.12]	--	1.528 [0.07]

The mean [SD] over subjects is shown. The algorithm with the best performance for each metric at each SNR is highlighted in bold. OMP and EM-OMP have no spurious activity by design since they estimate all the components using latency windows before 300 ms.

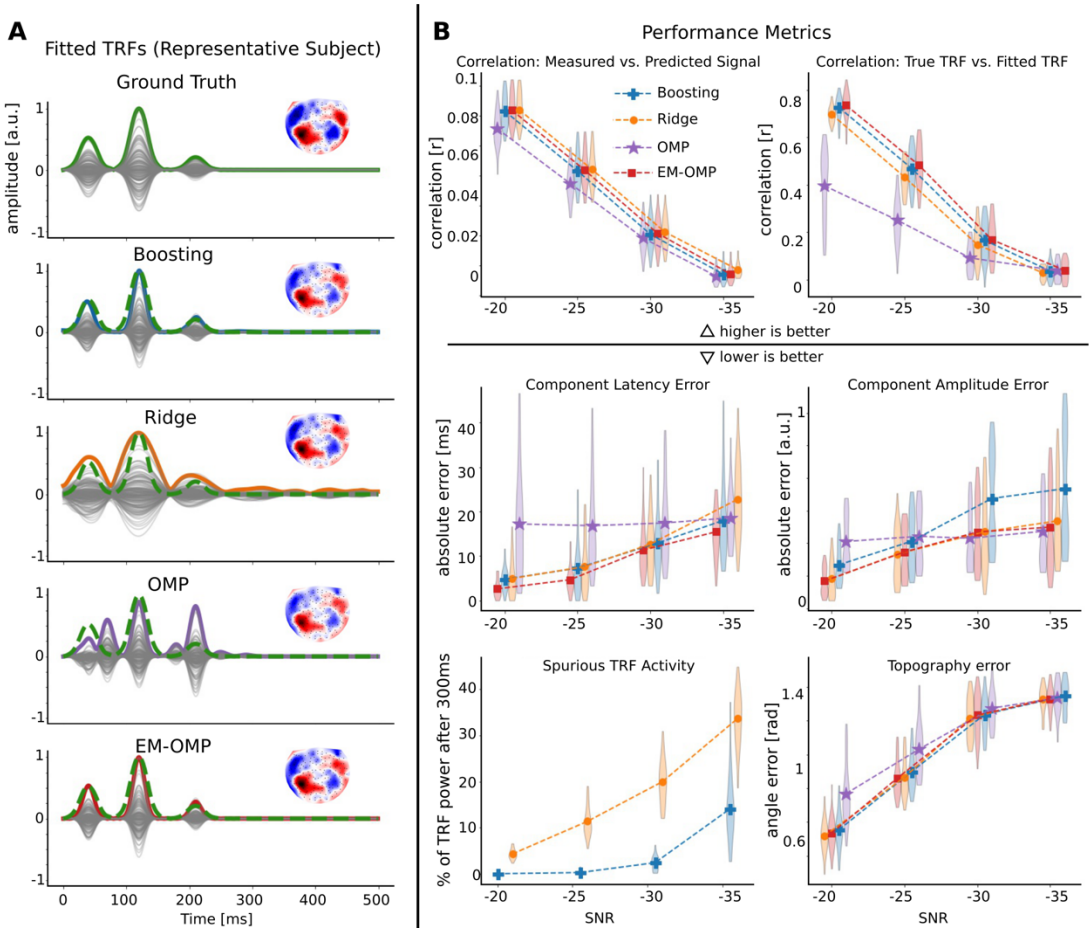

Figure 5.3. Performance comparison for DSS simulations. A. The fitted TRFs for a representative subject are shown, similar to the previous figure. The TRFs were fit on the first 6 DSS components and then back-projected to sensor space. All the algorithms except OMP result in reasonable TRF components and sensor topographies. **B.** Algorithm comparison using the performance metrics, similar to those shown in the previous figure. All the algorithms except OMP perform comparably, while the latter performs the worst in most cases.

5.4.4. Simulation: Source Space TRFs

Source space simulations were constructed with dipoles in auditory areas for each TRF component. These dipoles were projected onto sensor space using the forward model and source localized back to source space in order to simulate source localized MEG data. The algorithms were fit on these source localized signals and performance was compared using the same metrics (see Table 5.4, Fig. 5.4). Results were similar to the sensor space simulation, with EM-OMP outperforming the others and ridge regression and boosting giving comparable results (with ridge regression typically marginally better than boosting for most measures except spurious activity).

Table 5.4. Performance comparison for source space simulations

SNR	Algorithm	Correlation (Pearson r)	TRF correlation (Pearson r)	Latency error (m.s.)	Amplitude error (a.u.)	Spurious activity (% power > 300ms)	Source distribution error (MSE)
-15	Boosting	0.163 [0.02]	0.733 [0.03]	7.33 [3.1]	0.145 [0.07]	**4.635 [1.6]**	0.751 [0.2]
	Ridge	0.167 [0.02]	0.701 [0.02]	3.83 [2.3]	0.121 [0.04]	12.921 [2.1]	0.401 [0.1]
	OMP	0.153 [0.01]	0.540 [0.03]	10.44 [6.4]	0.126 [0.03]	--	0.991 [0.2]
	EM-OMP	**0.173 [0.02]**	**0.86 [0.03]**	**2.74 [2.1]**	**0.029 [0.01]**	--	**0.261 [0.1]**
-20	Boosting	0.079 [0.01]	0.574 [0.03]	9.94 [5.1]	0.165 [0.06]	**12.728 [2.9]**	0.935 [0.2]
	Ridge	0.086 [0.01]	0.576 [0.03]	9.05 [6.9]	0.128 [0.05]	20.722 [2.1]	0.552 [0.2]
	OMP	0.080 [0.01]	0.462 [0.04]	11.24 [5.0]	0.131 [0.04]	--	1.050 [0.2]
	EM-OMP	**0.092 [0.01]**	**0.794 [0.03]**	**1.95 [1.8]**	**0.050 [0.01]**	--	**0.481 [0.2]**
-25	Boosting	0.035 [0.007]	0.384 [0.04]	12.22 [6.1]	0.135 [0.08]	**23.224 [3.0]**	1.216 [0.2]
	Ridge	0.042 [0.008]	0.429 [0.04]	9.61 [3.9]	0.171 [0.08]	28.657 [2.1]	**0.790 [0.2]**
	OMP	0.039 [0.007]	0.354 [0.04]	11.51 [3.8]	0.145 [0.05]	--	1.197 [0.2]
	EM-OMP	**0.046 [0.008]**	**0.642 [0.06]**	**2.33 [1.7]**	**0.090 [0.03]**	--	0.808 [0.2]
-30	Boosting	0.016 [0.004]	0.238 [0.03]	21.7 [7.5]	**0.132 [0.05]**	**31.658 [3.4]**	1.456 [0.1]
	Ridge	**0.023 [0.004]**	0.297 [0.03]	20.02 [6.4]	0.201 [0.09]	34.589 [2.3]	1.219 [0.2]
	OMP	0.019 [0.004]	0.242 [0.03]	13.66 [4.9]	0.233 [0.09]	--	1.375 [0.1]
	EM-OMP	0.022 [0.005]	**0.478 [0.06]**	**3.33 [2.0]**	0.173 [0.04]	--	**1.112 [0.2]**

The mean [SD] over subjects is shown. The algorithm with the best performance for each metric at each SNR is highlighted in bold. OMP and EM-OMP have no spurious activity by design since they estimate all the components using latency windows before 300 ms.

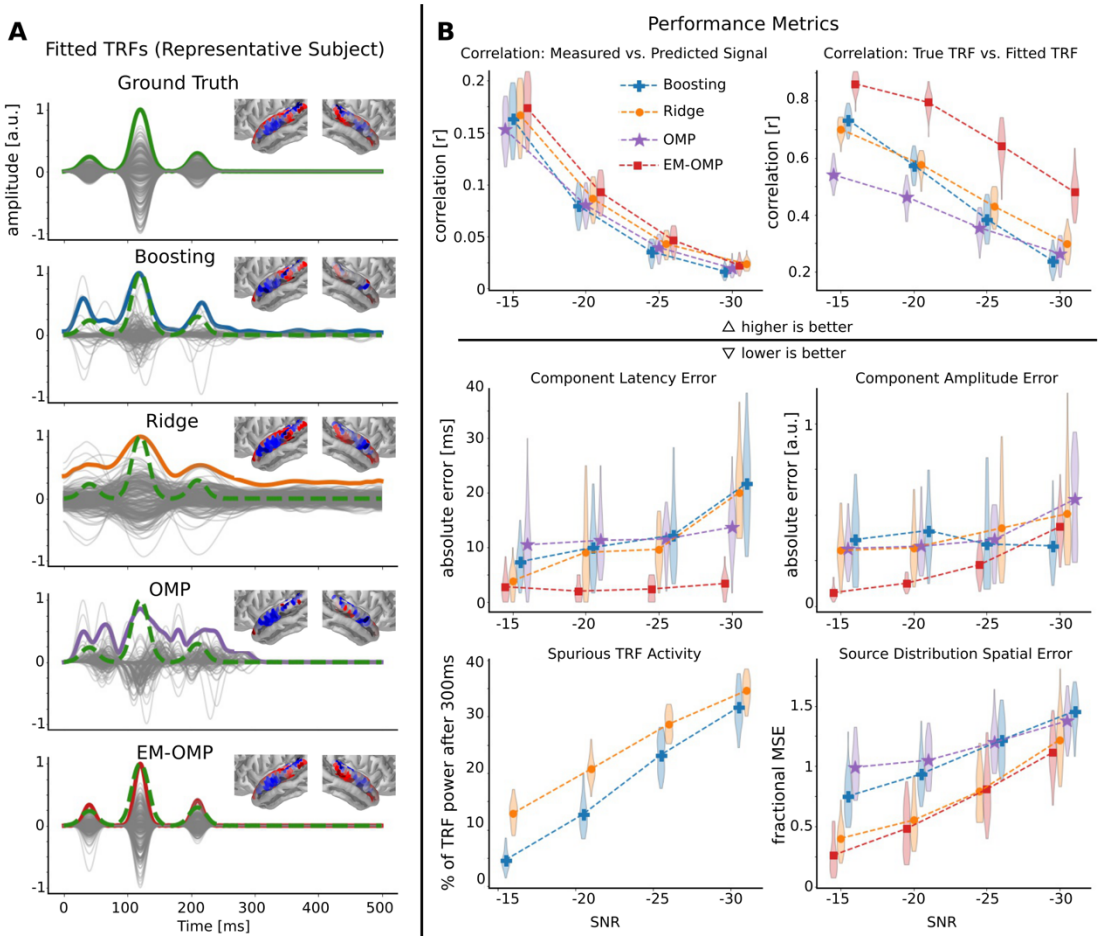

Figure 5.4. Performance comparison for source space simulations. A. The fitted TRFs for a representative subject are shown, similar to the previous figure. The source distributions in the temporal ROI at the largest peak near 100 ms are shown as insets. Boosting and OMP result in much sparser source distributions, and all the algorithms except OMP perform comparably in estimating the TRF components, although the ridge regression TRF has a lot more activity that may make it difficult to interpret in realistic situations where the ground truth is unknown. **B.** Algorithm comparison using the performance metrics, similar to those shown in the previous figure. EM-OMP outperforms the others in most cases.

5.4.5. Performance on real MEG data

The algorithms were compared on a real MEG dataset collected for a cocktail party experiment. Sensor space, DSS and source space TRFs are shown in Fig. 5.5. The only metric used was the correlation between the measured and predicted signals, since the other metrics cannot be calculated when the ground truth TRF components are unknown. The correlation mean and standard deviation are given in Table 5.5. Interestingly, ridge regression performs the best in terms of correlation, while the other three algorithms give marginally lower results. However, it is unclear if correlation is the most suitable metric for evaluating the accuracy of estimating TRF components. The individual ridge regression TRFs show a lot of activity and are harder to interpret than the boosting and EM-OMP TRFs, while the boosting TRFs seem overly sparse in some cases (see the sensor topographies in Fig. 5.5A).

Table 5.5. Correlation between measured and predicted signals for real data

Algorithm	Sensor Space	DSS	Source Space
Boosting	0.019 [0.014]	0.089 [0.027]	0.059 [0.031]
Ridge	**0.028 [0.014]**	**0.098 [0.027]**	**0.074 [0.032]**
OMP	0.021 [0.012]	0.080 [0.023]	0.063 [0.027]
EM-OMP	0.023 [0.013]	0.081 [0.024]	0.060 [0.028]

The correlation mean [SD] across subjects is shown, with the algorithm with the highest value for each case highlighted in bold.

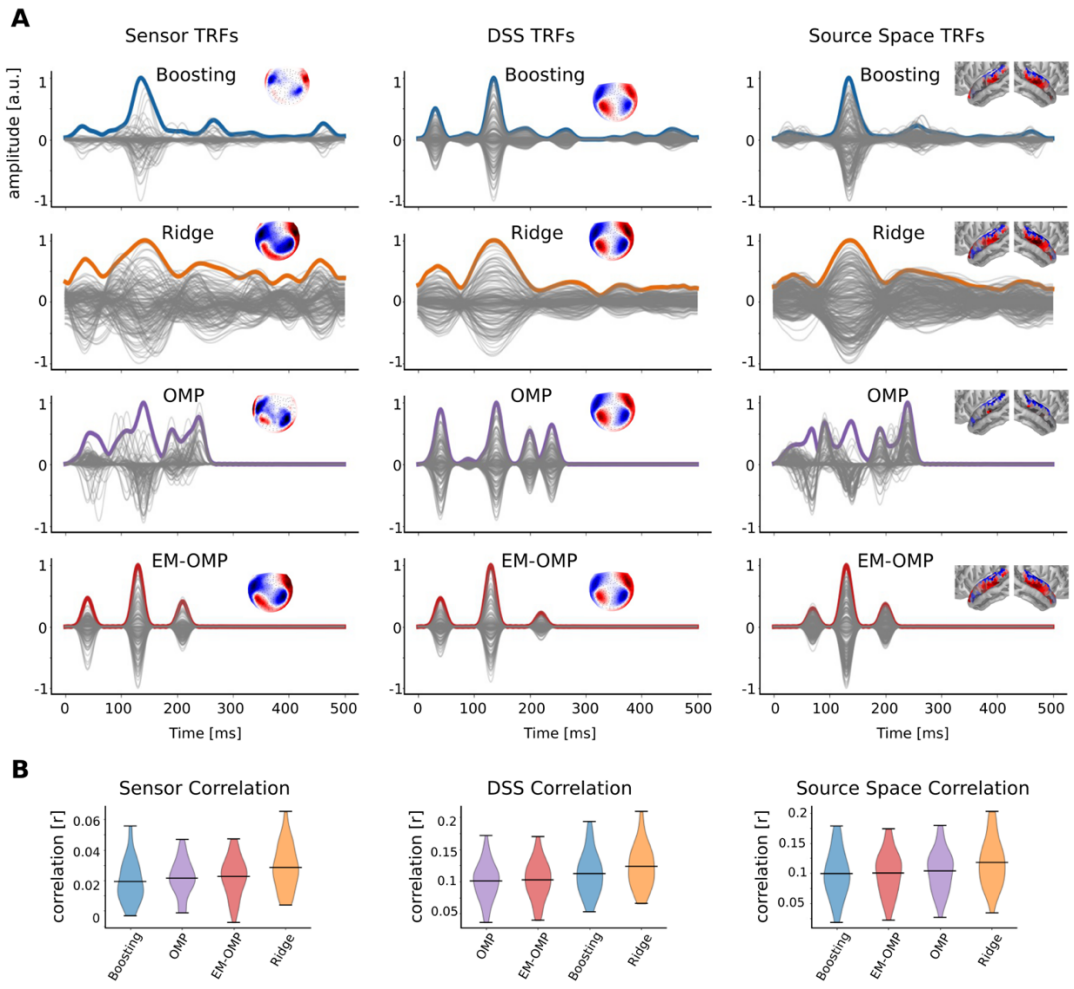

Figure 5.5. Performance comparison on real MEG data. A. The estimated sensor, DSS and source localized TRFs are shown for a representative subject. The sensor topographies and source distributions at the largest peak around 120 ms are shown as insets. The DSS and source localized TRFs are much cleaner than the sensor TRFs for both boosting and ridge regression. All the algorithms except OMP estimate similar TRF components and topographies. Note that the sensor space EM-OMP TRF has clear components and topographies, unlike the boosting TRF with overly sparse topographies or the ridge regression TRF with a lot of hard to interpret activity. However, boosting, ridge regression and EM-OMP show clear components and spatial patterns for the DSS and source localized TRFs. **B.** Correlation between the measured and predicted signals is shown as

a measure of model fit. Violin plots across subjects are shown for each algorithm in ascending order of their mean from left to right. Ridge regression consistently gives the highest correlations in all cases.

5.5. Discussion

The TRF framework has allowed auditory experiments to move away from trial averaged responses to repetitive stimuli, towards more naturalistic speech paradigms and has led to remarkable insights into the mechanisms underlying cortical processing of continuous speech. Despite significant advancements in experimental designs, it is unclear whether the specific algorithms employed for TRF estimation could bias model results. Prior work has compared variations of regularized regression and machine learning methods for linear models, in terms of their ability to decode subject attention in a multi-talker scenario based on prediction accuracy (Crosse et al., 2021; Geirnaert et al., 2021). However, several insights into neural processing of speech have arisen not only from the overall prediction accuracy of TRF models, but also from the specific characteristics of TRF components (Brodbeck et al., 2020b, 2018b, 2018a; Broderick et al., 2018).

In this work, we compared two commonly used TRF estimation algorithms, boosting and ridge regression, in terms of their ability to estimate these TRF components. Additionally, we proposed two algorithms based on OMP and EM that directly estimate these components. The OMP algorithm has been used extensively for sparse signal recovery (Tropp and Gilbert, 2007) and is typically capable of recovering

components in an efficient manner. The EM algorithm is a maximum likelihood method that is able to incorporate 'hidden' variables and is widely used in signal estimation (Do and Batzoglou, 2008). Both matching pursuit and EM have been used for single trial evoked response estimation (Limpiti et al., 2010; Sieluzycki et al., 2009), and here, we employ natural extensions of these algorithms for TRF component estimation. We discuss the performance of each algorithm on both simulations and real data in the following sections.

5.5.1. Performance in estimating TRFs as measured by correlation

The conventional measure for performance of TRF models is the correlation between the actual and the predicted signals. In order to avoid unreasonably high correlations due to overfitting, it is essential to calculate this correlation using test data that was not used for fitting the TRF models. In this work we used a nested cross-validation procedure to reduce overfitting from the estimated correlations for all four algorithms. The simulation results indicated that both boosting and ridge regression are comparable in terms of correlation, with ridge regression typically performing slightly better. Interestingly, OMP has higher correlation than ridge regression and boosting in the high noise single channel simulations, while EM-OMP outperforms the others by a large margin in the sensor and source space simulations. OMP and EM-OMP do not show any improvement over ridge regression and boosting for the DSS simulation. These results indicate that OMP and EM-OMP are suitable for estimating TRFs in high noise conditions, assuming that the appropriate latency windows can be determined a-

priori. Additionally, a suitable denoising technique such as DSS can result in ridge regression and boosting having correlation values comparable to EM-OMP.

However, correlation between the actual and the predicted signals may not always be an appropriate measure of TRF component estimation, since it depends on a variety of factors including SNR and predictor characteristics. High correlations may also result from overfitting and this metric would not penalize time-shift errors or spurious activity in the TRF. In light of this, we used several other metrics that directly measure the ability of these algorithms to estimate TRF components.

5.5.2. Performance in estimating TRF components

Many neurophysiological studies are primarily interested in specific TRF components (e.g., the M50, M100 and M200 which are analogous to the P1, N1, and P2 components of an auditory evoked response) and possible group or task differences in component amplitudes and latencies, rather than the entire TRF. Hence, we used simulated TRFs with known component latencies and amplitudes to evaluate these algorithms. The OMP algorithm performs the best when estimating TRF components for single channel data, and the EM-OMP algorithm outperforms the others for sensor and source space TRFs. However, it should be noted that the component windows used for the simulation were identical to the component windows provided a-priori to these algorithms, which may explain their better performance. Latency and amplitude estimation was comparable for boosting and ridge regression, with the latter having marginally lower errors. Ridge regression also had lower spatial error compared to

boosting (sensor topography and source distribution errors), which may be due to the fact that a sparse estimation technique like boosting cannot capture smooth spatial patterns as well as ridge regression. However, after applying the DSS algorithm, ridge regression, boosting and EM-OMP once again showed comparable performance, highlighting the importance of denoising methods such as DSS when estimating TRFs from noisy multidimensional data. Spurious peaks after 300 ms, present in both ridge regression and boosting TRFs, could lead to difficulties in interpretation and to false positives when detecting TRF components in real data. Ridge regression TRF estimates had much larger amounts of spurious activity than boosting.

5.5.3. Performance on real data

The algorithms were also compared on real MEG data collected during a cocktail party experiment. Ridge regression performed better in terms of correlation, and the other three algorithms had comparable performance (see Fig 5.5B). Although ridge regression had the best correlation, as observed above, this does not immediately imply that the ridge regression TRFs provided the best component estimates. The correlation values were distributed over a large range across subjects, possibly indicating a high degree of inter-subject variability in neural SNR for time-locked responses. Ridge regression resulted in smooth TRFs with several peaks and large amounts of non-zero activity which made them more difficult to interpret, especially for the sensor and source space TRFs. Boosting, though performing poorly in terms of correlation, allowed for sparser TRFs with fewer peaks that were easier to interpret. EM-OMP was

restricted to finding only three TRF components, using fixed a-priori component windows. The fact that the EM-OMP algorithm may have performed worse than ridge regression for real data, even though it outperformed the others in the simulations, indicates that these a-priori component windows may not be suitable for all subjects. Indeed, the assumptions underlying the EM-OMP algorithm may not be suitable for a variety of reasons including; large amounts of individual variability in TRF component latencies, missing TRF components due to anatomical or functional differences, and individual variability in component waveforms and peak widths. However, even with these constraints, EM-OMP was often able to recover TRF components and spatial patterns comparable to ridge regression. In any case, post-hoc analysis of TRF components estimated using conventional algorithms is also typically performed under similar assumptions (i.e., detecting TRF peaks using similar latency windows). Additionally, the ridge regression peaks were much broader than the EM-OMP peaks, suggesting that the latter may have suffered due to fixed and narrow waveforms in the basis dictionary. Therefore, without knowledge of the ground truth, it is difficult to judge which algorithm is most suitable for estimating TRF components.

5.5.4. Extensions and Applications

Careful tuning of the regularization parameter may improve the performance of ridge regression, at the cost of additional computational time. Variations on regularized regression, such as Lasso and Elastic Net, may also provide improvements in TRF estimation (Wong et al., 2018). The a-priori component windows used for OMP and

EM-OMP may need to be tuned for each predictor type or experiment, and possibly on a per-subject basis. The EM-OMP algorithm is also sensitive to initialization, and in our case was initialized using the components extracted from the OMP TRF. Extensions to EM-OMP based on dictionary learning may be able to directly estimate the component waveform instead of assuming a fixed shape using a basis dictionary.

Modern TRF studies use multiple types of predictors, each of which may impact the performance of these algorithms (e.g., continuous envelopes and impulse predictors to denote phoneme onsets; see Brodbeck et al., 2018a; Broderick et al., 2018; Di Liberto et al., 2015). For these experiments, banded ridge regression, which estimates different regularization parameters for each type of predictor, may improve performance over conventional ridge regression (Crosse et al., 2021). However, more basic science is needed before the OMP and EM-OMP algorithm can be applied for these cases, since the appropriate component latency windows of TRFs to higher level predictors must be determined.

The TRF framework has also been used to decode attention from neural responses during a cocktail party paradigm. Prior work has compared algorithms for estimating both TRFs (forward models) and decoders (backward models), in terms of their performance in attention decoding (Geirnaert et al., 2021; Wong et al., 2018). These studies used variations of regularized regression or machine learning methods, and it is unclear how they compare to sparse estimation techniques such as boosting, OMP or EM-OMP. Furthermore, attention decoding using forward models is typically performed by comparing the correlation values of the foreground and background

TRFs. However, as previously discussed, a higher correlation value (and better attention decoding) may not necessarily be the best measure for studies interested in accurate estimation of TRF components.

5.5.5. Conclusion

In this work, we compared the commonly used ridge regression and boosting algorithms and the novel OMP and EM-OMP algorithms in terms of their ability to estimate TRF component amplitudes and latencies. EM-OMP performed the best in the simulations, but perhaps underperformed on the real data, possibly because the a-priori assumptions on component latencies were not suitable. Boosting and ridge regression were comparable in terms of model fit and estimation errors in the simulations. Interestingly, for the real data, ridge regression resulted in higher correlation between the actual and predicted signals. However, in general, ridge regression TRFs displayed more spurious activity, while boosting resulted in more interpretable sparse TRFs. Our results indicate that EM-OMP may only perform well if its a-priori assumptions are realistic, while both ridge regression and boosting perform comparably in most cases.

Chapter 6

Conclusion

Speech is comprised of complex continuous signals that contain multiple levels of information, ranging from acoustics to words to sentences. This dissertation provides several insights into the neural mechanisms underlying our ability to perceive and comprehend speech. It showcases the use of MEG experiments along with TRFs for investigating cortical responses to continuous speech. In this chapter, the strengths and weaknesses of the TRF framework are discussed, followed by a summary of the main results of this dissertation and a discussion of possible avenues for future research.

6.1. The TRF model: Advantages and Disadvantages

TRF models with multiple predictors allow for simultaneous investigation of several levels of speech processing. In this work, both low level speech processing and high-level word and sentence processing of continuous speech were investigated using TRFs. Additionally, neural tracking of speech features, along with the distinct cortical networks involved in processing different aspects of speech (such as spoken equations vs. sentences) were explored. Finally, TRFs were used in conjunction with the cocktail party paradigm, to explore attentional mechanisms involved in speech comprehension.

However, it should be noted that TRFs are limited in several aspects. Firstly, M/EEG TRF models detect aggregate activity of large populations of neurons involved in speech processing, and may not be able to detect fine-grained details about the underlying neuronal mechanisms. Secondly, TRFs are only able to model time-locked activity, even though cortical processing of speech may not be strictly time-locked to speech features. This issue may be especially prevalent for TRFs to high-level speech features, for two reasons; high-level processing is likely less time-locked, and the exact timing of high-level features may be ambiguous (e.g., the timing of impulses denoting word onsets). Finally, TRFs are a linear model and cannot capture all the complexities of the nonlinear systems involved in speech processing. Although increasingly complex nonlinear representations of speech (e.g., envelopes, envelope onsets, phoneme surprisal, word onsets etc.) have been employed to mitigate this problem, nonlinear methods such as neural networks may prove to be more effective in modelling these nonlinearities. However, even if these nonlinear methods result in better model fits, they often involve difficult-to-interpret models and may not provide meaningful insights into the underlying cortical mechanisms. In contrast, TRFs are typically comprised of informative waveforms with distinct and localized components arising from specific cortical processes.

Despite these concerns, the TRF framework is capable of providing robust and interpretable models of time-locked processing of continuous stimuli and has led to several insights into speech processing (Brodbeck and Simon, 2020). The many advantages of TRFs are showcased in several of the results reported in this dissertation.

6.2. Summary of main results

Chapter 3 of this dissertation investigated time-locked responses to continuous speech in the high gamma range and was inspired by studies into cortical frequency following responses. This work resulted in several key advancements in the field of high frequency responses to speech. Firstly, high gamma time-locked cortical responses were detected to continuous speech, which is a more naturalistic stimuli than the repeated speech syllables used in conventional FFR studies. Secondly, a TRF model with multiple predictors revealed that these responses were predominantly to the high frequency envelope modulation of the speech signal. Thirdly, these responses were found to time-lock to the low pitch segments of speech. Interestingly, no age-related differences were detected in these responses. This work provides insights into low-level cortical processing of speech and bridges the gap between the well-known FFR and the low frequency envelope following TRF.

Next, chapter 4 explored cortical responses to spoken equations and sentences, using TRFs as well as frequency domain techniques and decoding methods. Firstly, this work showed that cortical tracking of sentences and equations is present only when subjects attend to the relevant speaker in a cocktail party paradigm. Secondly, responses from distinct cortical networks involved in sentence and equation processing were detected and these responses were correlated with performance in detecting equation and sentence deviants, indicating that they may be linked to calculation and comprehension. Next, TRF analysis using high-level sentence and equation onset predictors revealed the spatiotemporal dynamics of these cortical networks.

Furthermore, these cortical responses could also be used to decode whether subjects attended to sentences or equations. This work showcases the suitability of cocktail party paradigms used in conjunction with techniques such as TRFs to probe high-level cortical processing of speech, and provides insights into the neural mechanisms involved in arithmetic and language processing.

Finally, chapter 5 of this dissertation compared both conventional and novel TRF algorithms, in terms of their performance in estimating TRF components. The novel algorithms were based on a-priori assumptions about typical TRF components and directly estimated component latencies and amplitudes. Although these algorithms performed well in the simulations, their weak performance on real data may indicate that such assumptions may not account for individual variability in TRF components. The conventional algorithms performed comparably in most cases, with ridge regression resulting in higher correlation values. The results indicated that additional concerns such as overly sparse or spurious TRF activity must also be considered when selecting an appropriate algorithm. This work provides an initial investigation into the performance and biases of these algorithms and highlights key concerns when estimating and interpreting TRF components.

6.3. Future directions

The results of chapter 3 allow for several avenues of future exploration. Firstly, the lack of significant age-related differences in high gamma TRFs needs to be explored further. Some studies indicate that older adults may have weaker cortical FFR (Ross et

al., 2020), and experiments with larger subject populations listening to continuous speech are needed to obtain a conclusive result. Secondly, an intriguing question is whether the high gamma TRF would show attentional modulation, like the M100 component of the low frequency envelope TRF. Cocktail party experiments may shed light on this question. Thirdly, the neural origin of these high gamma responses is unclear. Some recent invasive studies have indicated that cortical FFRs arise from thalamorecepient layers of cortex (Gnanateja et al., 2021), in line with the hypotheses discussed in section 3.5.8 of this dissertation. Combined M/EEG experiments with simultaneous detection of subcortical responses by EEG and cortical responses by MEG may provide further insight into the neural origin of high gamma responses to continuous speech.

The work presented in chapter 4 demonstrates the suitability of the cocktail party paradigm and TRFs for investigating high-level processing of speech content, and could lead to several potential future directions. Firstly, more nuanced experimental designs may be able to tease apart the specific neural mechanisms involved in sentence and equation processing (e.g., distinguishing between responses arising from cortical processes involved in detecting equation boundaries, parsing equations, identifying the arithmetic operation, or computing the equation result). Secondly, these TRF methods could also be used for natural non-rhythmic cocktail party stimuli with sentences and equations, with careful construction of high-level predictors such as sentence and equation onsets. Such experiments may provide insights into whether the TRF peaks seen in this work arise from onset responses or processing of the completion of the

sentence. Finally, the isochronous cocktail party paradigm used in this work (which was pioneered by Ding and others to investigate aspects of language processing; see Ding et al., 2018) can also be employed for a wide variety of other stimuli and could be used to investigate cortical processing of high-level syntactic, semantic or logical structures in continuous speech.

The results from Chapter 5 indicate that further work is needed to determine whether algorithms such as OMP and EM-OMP would be suitable for real data given more realistic assumptions on TRF components. It may also be possible to improve performance by extending these algorithms using dictionary learning methods or Bayesian frameworks. Conventional algorithms such as boosting and ridge regression must also be compared using more complex TRF models since their performance may vary based on the type of predictors (e.g., impulse predictors such as word onsets). Finally, the performance of sparse algorithms such as boosting, OMP and EM-OMP for decoding attention in cocktail party paradigms must also be evaluated. Improvements in TRF estimation methods may lead to advancements in hearing aid technology and in understanding, diagnosing, and treating hearing and speech deficits.

Bibliography

Abd Hamid, A.I., Yusoff, A.N., Mukari, S.Z.-M.S., Mohamad, M., 2011. Brain Activation during Addition and Subtraction Tasks In-Noise and In-Quiet. Malays J Med Sci 18, 3–15.

Aiken, S.J., Picton, T.W., 2008. Envelope and spectral frequency-following responses to vowel sounds. Hearing Research 245, 35–47. https://doi.org/10.1016/j.heares.2008.08.004

Akram, S., Presacco, A., Simon, J.Z., Shamma, S.A., Babadi, B., 2016. Robust decoding of selective auditory attention from MEG in a competing-speaker environment via state-space modeling. NeuroImage 124, 906–917. https://doi.org/10.1016/j.neuroimage.2015.09.048

Alain, C., Roye, A., Salloum, C., 2014. Effects of age-related hearing loss and background noise on neuromagnetic activity from auditory cortex. Front Syst Neurosci 8. https://doi.org/10.3389/fnsys.2014.00008

Amalric, M., Dehaene, S., 2019. A distinct cortical network for mathematical knowledge in the human brain. NeuroImage 189, 19–31. https://doi.org/10.1016/j.neuroimage.2019.01.001

Amalric, M., Dehaene, S., 2018. Cortical circuits for mathematical knowledge: evidence for a major subdivision within the brain's semantic networks. Philosophical Transactions of the Royal Society B: Biological Sciences 373, 20160515. https://doi.org/10.1098/rstb.2016.0515

Amalric, M., Dehaene, S., 2016. Origins of the brain networks for advanced mathematics in expert mathematicians. Proceedings of the National Academy of Sciences 113, 4909–4917. https://doi.org/10.1073/pnas.1603205113

Anderson, S., Parbery-Clark, A., White-Schwoch, T., Kraus, N., 2012. Aging Affects Neural Precision of Speech Encoding. Journal of Neuroscience 32, 14156–14164. https://doi.org/10.1523/JNEUROSCI.2176-12.2012

Arsalidou, M., Taylor, M.J., 2011. Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. NeuroImage 54, 2382–2393. https://doi.org/10.1016/j.neuroimage.2010.10.009

Assmann, P., Summerfield, Q., 2004. The Perception of Speech Under Adverse Conditions, in: Greenberg, S., Ainsworth, W.A., Popper, A.N., Fay, R.R.

(Eds.), Speech Processing in the Auditory System, Springer Handbook of Auditory Research. Springer, New York, NY, pp. 231–308. https://doi.org/10.1007/0-387-21575-1_5

Attal, Y., Bhattacharjee, M., Yelnik, J., Cottereau, B., Lefevre, J., Okada, Y., Bardinet, E., Chupin, M., Baillet, S., 2007. Modeling and Detecting Deep Brain Activity with MEG EEG, in: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Presented at the 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4937–4940. https://doi.org/10.1109/IEMBS.2007.4353448

Avancini, C., Soltész, F., Szűcs, D., 2015. Separating stages of arithmetic verification: An ERP study with a novel paradigm. Neuropsychologia 75, 322–329. https://doi.org/10.1016/j.neuropsychologia.2015.06.016

Baggio, G., Hagoort, P., 2011. The balance between memory and unification in semantics: A dynamic account of the N400. Language and Cognitive Processes 26, 1338–1367. https://doi.org/10.1080/01690965.2010.542671

Baillet, S., 2017. Magnetoencephalography for brain electrophysiology and imaging. Nature Neuroscience 20, 327–339. https://doi.org/10.1038/nn.4504

Balderston, N.L., Schultz, D.H., Baillet, S., Helmstetter, F.J., 2014. Rapid Amygdala Responses during Trace Fear Conditioning without Awareness. PLOS ONE 9, e96803. https://doi.org/10.1371/journal.pone.0096803

Baldo, J.V., Dronkers, N.F., 2007. Neural correlates of arithmetic and language comprehension: A common substrate? Neuropsychologia 45, 229–235. https://doi.org/10.1016/j.neuropsychologia.2006.07.014

Barbati, G., Porcaro, C., Zappasodi, F., Rossini, P.M., Tecchio, F., 2004. Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals. Clinical Neurophysiology 115, 1220–1232. https://doi.org/10.1016/j.clinph.2003.12.015

Basu, M., Krishnan, A., Weber-Fox, C., 2010. Brainstem correlates of temporal auditory processing in children with specific language impairment: Brainstem correlates of temporal processing. Developmental Science 13, 77–91. https://doi.org/10.1111/j.1467-7687.2009.00849.x

Bemis, D.K., Pylkkänen, L., 2013. Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. Cereb. Cortex 23, 1859–1873. https://doi.org/10.1093/cercor/bhs170

Bidelman, G.M., 2018. Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. Neuroimage 175, 56–69. https://doi.org/10.1016/j.neuroimage.2018.03.060

Bidelman, G.M., 2015. Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR. Hearing Research 323, 68–80. https://doi.org/10.1016/j.heares.2015.01.011

Bidelman, G.M., Villafuerte, J.W., Moreno, S., Alain, C., 2014. Age-related changes in the subcortical–cortical encoding and categorical perception of speech. Neurobiology of Aging 35, 2526–2540. https://doi.org/10.1016/j.neurobiolaging.2014.05.006

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. Cereb Cortex 19, 2767–2796. https://doi.org/10.1093/cercor/bhp055

Blanco-Elorrieta, E., Ding, N., Pylkkänen, L., Poeppel, D., 2019. Understanding requires tracking: noise and knowledge interact in bilingual comprehension (preprint). Neuroscience. https://doi.org/10.1101/609628

Boersma, P., 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proceedings of the Institute of Phonetic Sciences 97–110.

Boersma, P., Weenick, D., 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.43 [WWW Document]. URL http://www.praat.org (accessed 9.8.18).

Brodbeck, C., Brooks, T.L., Das, P., Reddigari, S., 2019. christianbrodbeck/Eelbrain: 0.30. Zenodo. https://doi.org/10.5281/zenodo.2653785

Brodbeck, C., Das, P., jpkulasingham, Reddigari, S., Brooks, T.L., 2021a. Eelbrain 0.36. Zenodo. https://doi.org/10.5281/zenodo.5152554

Brodbeck, C., Das, P., Kulasingham, J.P., Bhattasali, S., Gaston, P., Resnik, P., Simon, J.Z., 2021b. Eelbrain: A Python toolkit for time-continuous analysis with temporal response functions. https://doi.org/10.1101/2021.08.01.454687

Brodbeck, C., Hong, L.E., Simon, J.Z., 2018a. Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech. Current Biology 28, 3976-3983.e5. https://doi.org/10.1016/j.cub.2018.10.042

Brodbeck, C., Jiao, A., Hong, L.E., Simon, J.Z., 2020a. Neural speech restoration at the cocktail party: Auditory cortex recovers masked speech of both attended and ignored speakers. PLOS Biology 18, e3000883. https://doi.org/10.1371/journal.pbio.3000883

Brodbeck, C., Jiao, A., Hong, L.E., Simon, J.Z., 2020b. Neural speech restoration at the cocktail party: Auditory cortex recovers masked speech of both attended and ignored speakers. bioRxiv 866749. https://doi.org/10.1101/866749

Brodbeck, C., Presacco, A., Anderson, S., Simon, J.Z., 2018b. Over-Representation of Speech in Older Adults Originates from Early Response in Higher Order Auditory Cortex. Acta Acustica united with Acustica 104, 774–777. https://doi.org/10.3813/AAA.919221

Brodbeck, C., Presacco, A., Simon, J.Z., 2018c. Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension. NeuroImage 172, 162–174. https://doi.org/10.1016/j.neuroimage.2018.01.042

Brodbeck, C., Simon, J.Z., 2020. Continuous speech processing. Current Opinion in Physiology 18, 25–31. https://doi.org/10.1016/j.cophys.2020.07.014

Broderick, M.P., Anderson, A.J., Di Liberto, G.M., Crosse, M.J., Lalor, E.C., 2018. Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. Current Biology 28, 803-809.e3. https://doi.org/10.1016/j.cub.2018.01.080

Butler, B.E., Lomber, S.G., 2013. Functional and structural changes throughout the auditory system following congenital and early-onset deafness: implications for hearing restoration. Frontiers in Systems Neuroscience 7. https://doi.org/10.3389/fnsys.2013.00092

Cai, T.T., Wang, L., 2011. Orthogonal Matching Pursuit for Sparse Signal Recovery With Noise. IEEE Transactions on Information Theory 57, 4680–4688. https://doi.org/10.1109/TIT.2011.2146090

Campbell, J.I.D., Austin, S., 2002. Effects of response time deadlines on adults' strategy choices for simple addition. Memory & Cognition 30, 988–994. https://doi.org/10.3758/BF03195782

Caspary, D.M., Ling, L., Turner, J.G., Hughes, L.F., 2008. Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. Journal of Experimental Biology 211, 1781–1791. https://doi.org/10.1242/jeb.013581

Caspary, D.M., Llano, D.A., 2019. Aging Processes in the Subcortical Auditory System, in: Kandler, K. (Ed.), The Oxford Handbook of the Auditory Brainstem. Oxford University Press, pp. 638–680. https://doi.org/10.1093/oxfordhb/9780190849061.013.16

Castaldi, E., Piazza, M., Dehaene, S., Vignaud, A., Eger, E., 2019. Attentional amplification of neural codes for number independent of other quantities along the dorsal visual stream. eLife 8, e45160. https://doi.org/10.7554/eLife.45160

Cha, K., Zatorre, R.J., Schönwiesner, M., 2016. Frequency Selectivity of Voxel-by-Voxel Functional Connectivity in Human Auditory Cortex. Cereb Cortex 26, 211–224. https://doi.org/10.1093/cercor/bhu193

Chambers, A.R., Resnik, J., Yuan, Y., Whitton, J.P., Edge, A.S., Liberman, M.C., Polley, D.B., 2016. Central Gain Restores Auditory Processing following Near-Complete Cochlear Denervation. Neuron 89, 867–879. https://doi.org/10.1016/j.neuron.2015.12.041

Chen, Y., Jin, P., Ding, N., 2020. The influence of linguistic information on cortical tracking of words. Neuropsychologia 107640. https://doi.org/10.1016/j.neuropsychologia.2020.107640

Cherry, E.C., 1953. Some Experiments on the Recognition of Speech, with One and with Two Ears. The Journal of the Acoustical Society of America 25, 975–979. https://doi.org/10.1121/1.1907229

Cheveigné, A. de, Simon, J.Z., 2008. Denoising based on spatial filtering. J Neurosci Methods 171, 331–339. https://doi.org/10.1016/j.jneumeth.2008.03.015

Chittka, L., Brockmann, A., 2005. Perception Space—The Final Frontier. PLOS Biology 3, e137. https://doi.org/10.1371/journal.pbio.0030137

Christian Brodbeck, Teon L Brooks, Proloy Das, Samir Reddigari, Joshua P Kulasingham, 2020. christianbrodbeck/Eelbrain: 0.33. Zenodo. https://doi.org/10.5281/zenodo.4060224

Coffey, E.B.J., Chepesiuk, A.M.P., Herholz, S.C., Baillet, S., Zatorre, R.J., 2017a. Neural Correlates of Early Sound Encoding and their Relationship to Speech-

in-Noise Perception. Front Neurosci 11, 479. https://doi.org/10.3389/fnins.2017.00479

Coffey, E.B.J., Herholz, S.C., Chepesiuk, A.M.P., Baillet, S., Zatorre, R.J., 2016. Cortical contributions to the auditory frequency-following response revealed by MEG. Nature Communications 7, 11070. https://doi.org/10.1038/ncomms11070

Coffey, E.B.J., Musacchia, G., Zatorre, R.J., 2017b. Cortical Correlates of the Auditory Frequency-Following and Onset Responses: EEG and fMRI Evidence. J. Neurosci. 37, 830–838. https://doi.org/10.1523/JNEUROSCI.1265-16.2016

Cornwell, B.R., Arkin, N., Overstreet, C., Carver, F.W., Grillon, C., 2012. Distinct contributions of human hippocampal theta to spatial cognition and anxiety. Hippocampus 22, 1848–1859. https://doi.org/10.1002/hipo.22019

Cornwell, B.R., Carver, F.W., Coppola, R., Johnson, L., Alvarez, R., Grillon, C., 2008. Evoked amygdala responses to negative faces revealed by adaptive MEG beamformers. Brain Research 1244, 103–112. https://doi.org/10.1016/j.brainres.2008.09.068

Crosse, M.J., Di Liberto, G.M., Bednar, A., Lalor, E.C., 2016. The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. Front. Hum. Neurosci. 0. https://doi.org/10.3389/fnhum.2016.00604

Crosse, M.J., Zuk, N.J., Di Liberto, G.M., Nidiffer, A., Molholm, S., Lalor, E., 2021. Linear Modeling of Neurophysiological Responses to Naturalistic Stimuli: Methodological Considerations for Applied Research (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/jbz2w

Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., 2000. Dynamic Statistical Parametric Mapping: Combining fMRI and MEG for High-Resolution Imaging of Cortical Activity. Neuron 26, 55–67. https://doi.org/10.1016/S0896-6273(00)81138-1

Dastjerdi, M., Ozker, M., Foster, B.L., Rangarajan, V., Parvizi, J., 2013. Numerical processing in the human parietal cortex during experimental and natural conditions. Nature Communications 4, 2528. https://doi.org/10.1038/ncomms3528

David, S.V., Mesgarani, N., Shamma, S.A., 2007. Estimating sparse spectro-temporal receptive fields with natural stimuli. Network 18, 191–212. https://doi.org/10.1080/09548980701609235

de Cheveigné, A., Simon, J.Z., 2008. Sensor noise suppression. Journal of Neuroscience Methods 168, 195–202. https://doi.org/10.1016/j.jneumeth.2007.09.012

de Cheveigné, A., Simon, J.Z., 2007. Denoising based on Time-Shift PCA. J Neurosci Methods 165, 297–305. https://doi.org/10.1016/j.jneumeth.2007.06.003

de Munck, J.C., Bijma, F., Gaura, P., Sieluzycki, C.A., Branco, M.I., Heethaar, R.M., 2004. A maximum-likelihood estimator for trial-to-trial variations in noisy MEG/EEG data sets. IEEE Transactions on Biomedical Engineering 51, 2123–2128. https://doi.org/10.1109/TBME.2004.836515

Decruy, L., Vanthornhout, J., Francart, T., 2019. Evidence for enhanced neural tracking of the speech envelope underlying age-related speech-in-noise difficulties. Journal of Neurophysiology 122, 601–615. https://doi.org/10.1152/jn.00687.2018

Dehaene, S., 1999. Sources of Mathematical Thinking: Behavioral and Brain-Imaging Evidence. Science 284, 970–974. https://doi.org/10.1126/science.284.5416.970

Dehaene, S., Cohen, L., 1997. Cerebral Pathways for Calculation: Double Dissociation between Rote Verbal and Quantitative Knowledge of Arithmetic. Cortex 33, 219–250. https://doi.org/10.1016/S0010-9452(08)70002-9

Dehaene, S., Molko, N., Cohen, L., Wilson, A.J., 2004. Arithmetic and the brain. Current Opinion in Neurobiology 14, 218–224. https://doi.org/10.1016/j.conb.2004.03.008

Dehaene, S., Piazza, M., Pinel, P., Cohen, L., 2003. Three parietal circuits for number processing. Cognitive neuropsychology 20, 487–506. https://doi.org/10.1080/02643290244000239

Delazer, M., Girelli, L., Semenza, C., Denes, G., 1999. Numerical skills and aphasia. Journal of the International Neuropsychological Society 5, 213–221. https://doi.org/10.1017/S1355617799533043

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum Likelihood from Incomplete Data Via the EM Algorithm. Journal of the Royal Statistical

Society: Series B (Methodological) 39, 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Di Liberto, G.M., O'Sullivan, J.A., Lalor, E.C., 2015. Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Current Biology 25, 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030

Ding, N., Chatterjee, M., Simon, J.Z., 2014. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. NeuroImage 88, 41–46. https://doi.org/10.1016/j.neuroimage.2013.10.054

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., Poeppel, D., 2017. Characterizing Neural Entrainment to Hierarchical Linguistic Units using Electroencephalography (EEG). Front Hum Neurosci 11. https://doi.org/10.3389/fnhum.2017.00481

Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nature Neuroscience 19, 158–164. https://doi.org/10.1038/nn.4186

Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., Zhang, J., 2018. Attention Is Required for Knowledge-Based Sequential Grouping: Insights from the Integration of Syllables into Words. J. Neurosci. 38, 1178–1188. https://doi.org/10.1523/JNEUROSCI.2606-17.2017

Ding, N., Simon, J.Z., 2013. Adaptive Temporal Encoding Leads to a Background-Insensitive Cortical Representation of Speech. J. Neurosci. 33, 5728–5735. https://doi.org/10.1523/JNEUROSCI.5297-12.2013

Ding, N., Simon, J.Z., 2012a. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. Journal of Neurophysiology 107, 78–89. https://doi.org/10.1152/jn.00297.2011

Ding, N., Simon, J.Z., 2012b. Emergence of neural encoding of auditory objects while listening to competing speakers. PNAS 109, 11854–11859. https://doi.org/10.1073/pnas.1205381109

Do, C.B., Batzoglou, S., 2008. What is the expectation maximization algorithm? Nat Biotechnol 26, 897–899. https://doi.org/10.1038/nbt1406

Dumas, T., Dubal, S., Attal, Y., Chupin, M., Jouvent, R., Morel, S., George, N., 2013. MEG Evidence for Dynamic Amygdala Modulations by Gaze and Facial Emotions. PLoS One 8. https://doi.org/10.1371/journal.pone.0074145

Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., Kleinschmidt, A., 2009. Deciphering Cortical Number Coding from Human Brain Activity Patterns. Current Biology 19, 1608–1615. https://doi.org/10.1016/j.cub.2009.08.047

Eger, E., Sterzer, P., Russ, M.O., Giraud, A.-L., Kleinschmidt, A., 2003. A supramodal number representation in human intraparietal cortex. Neuron 37, 719–725. https://doi.org/10.1016/s0896-6273(03)00036-9

Elhilali, M., Fritz, J.B., Klein, D.J., Simon, J.Z., Shamma, S.A., 2004. Dynamics of Precise Spike Timing in Primary Auditory Cortex. J. Neurosci. 24, 1159–1172. https://doi.org/10.1523/JNEUROSCI.3825-03.2004

Fehr, T., Code, C., Herrmann, M., 2007. Common brain regions underlying different arithmetic operations as revealed by conjunct fMRI–BOLD activation. Brain Research 1172, 93–102. https://doi.org/10.1016/j.brainres.2007.07.043

Fiedler, L., Wöstmann, M., Herbst, S.K., Obleser, J., 2019. Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. NeuroImage 186, 33–42. https://doi.org/10.1016/j.neuroimage.2018.10.057

Fischl, B., 2012. FreeSurfer. NeuroImage 62, 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Forte, A.E., Etard, O., Reichenbach, T., 2017. The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. eLife 6, e27203. https://doi.org/10.7554/eLife.27203

Friederici, A.D., 2011. The Brain Basis of Language Processing: From Structure to Function. Physiological Reviews 91, 1357–1392. https://doi.org/10.1152/physrev.00006.2011

Friederici, A.D., 2002. Towards a neural basis of auditory sentence processing. Trends in Cognitive Sciences 6, 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

Geirnaert, S., Vandecappelle, S., Alickovic, E., de Cheveigne, A., Lalor, E., Meyer, B.T., Miran, S., Francart, T., Bertrand, A., 2021. Electroencephalography-Based Auditory Attention Decoding: Toward Neurosteered Hearing Devices. IEEE Signal Processing Magazine 38, 89–102. https://doi.org/10.1109/MSP.2021.3075932

Gelman, R., Butterworth, B., 2005. Number and language: how are they related? Trends in Cognitive Sciences 9, 6–10. https://doi.org/10.1016/j.tics.2004.11.004

Gnanateja, G.N., Rupp, K., Llanos, F., Remick, M., Pernia, M., Sadagopan, S., Teichert, T., Abel, T., Chandrasekaran, B., 2021. Deconstructing the Cortical Sources of Frequency Following Responses to Speech: A Cross-species Approach. https://doi.org/10.1101/2021.05.17.444462

Göbel, S., Walsh, V., Rushworth, M.F.S., 2001. The Mental Number Line and the Human Angular Gyrus. NeuroImage 14, 1278–1289. https://doi.org/10.1006/nimg.2001.0927

Goossens, T., Vercammen, C., Wouters, J., Wieringen, A. van, 2016. Aging Affects Neural Synchronization to Speech-Related Acoustic Modulations. Front. Aging Neurosci. 8. https://doi.org/10.3389/fnagi.2016.00133

Gordon-Salant, S., Yeni-Komshian, G.H., Fitzgibbons, P.J., Barrett, J., 2006. Age-related differences in identification and discrimination of temporal cues in speech segments. The Journal of the Acoustical Society of America 119, 2455–2466. https://doi.org/10.1121/1.2171527

Grabner, R.H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., Neuper, C., 2009. To retrieve or to calculate? Left angular gyrus mediates the retrieval of arithmetic facts during problem solving. Neuropsychologia 47, 604–608. https://doi.org/10.1016/j.neuropsychologia.2008.10.013

Grabner, R.H., Ansari, D., Reishofer, G., Stern, E., Ebner, F., Neuper, C., 2007. Individual differences in mathematical competence predict parietal brain activation during mental calculation. NeuroImage 38, 346–356. https://doi.org/10.1016/j.neuroimage.2007.07.041

Gramfort, A., 2013. MEG and EEG data analysis with MNE-Python. Frontiers in Neuroscience 7. https://doi.org/10.3389/fnins.2013.00267

Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M., 2014. MNE software for processing MEG and EEG data. Neuroimage 86, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Gramfort, Alexandre, Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M.S., 2014. MNE software for processing MEG and EEG data. NeuroImage 86, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Hämäläinen, M., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V., 1993. Magnetoencephalography---theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev. Mod. Phys. 65, 413–497. https://doi.org/10.1103/RevModPhys.65.413

Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med. Biol. Eng. Comput. 32, 35–42. https://doi.org/10.1007/BF02512476

Hansen, P.C., Kringelbach, M.L., Salmelin, R. (Eds.), 2010. MEG: an introduction to methods. Oxford University Press, New York.

Hartmann, T., Weisz, N., 2019. Auditory cortical generators of the Frequency Following Response are modulated by intermodal attention. NeuroImage 203, 116185. https://doi.org/10.1016/j.neuroimage.2019.116185

Harvey, B.M., Dumoulin, S.O., 2017. A network of topographic numerosity maps in human association cortex. Nature Human Behaviour 1, 1–9. https://doi.org/10.1038/s41562-016-0036

Harvey, B.M., Klein, B.P., Petridou, N., Dumoulin, S.O., 2013. Topographic Representation of Numerosity in the Human Parietal Cortex. Science 341, 1123–1126. https://doi.org/10.1126/science.1239052

He, N., Mills, J.H., Ahlstrom, J.B., Dubno, J.R., 2008. Age-related differences in the temporal modulation transfer function with pure-tone carriers. J Acoust Soc Am 124, 3841–3849. https://doi.org/10.1121/1.2998779

Herrmann, B., Henry, M.J., Johnsrude, I.S., Obleser, J., 2016. Altered temporal dynamics of neural adaptation in the aging human auditory cortex. Neurobiology of Aging 45, 10–22. https://doi.org/10.1016/j.neurobiolaging.2016.05.006

Hertrich, I., Dietrich, S., Trouvain, J., Moos, A., Ackermann, H., 2012. Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. Psychophysiology 49, 322–334. https://doi.org/10.1111/j.1469-8986.2011.01314.x

Hertrich, I., Mathiak, K., Lutzenberger, W., Ackermann, H., 2004. Transient and phase-locked evoked magnetic fields in response to periodic acoustic signals. Neuroreport 15, 1687–1690. https://doi.org/10.1097/01.wnr.0000134930.04561.b2

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nature Reviews Neuroscience 8, 393–402. https://doi.org/10.1038/nrn2113

Hill, K.T., Miller, L.M., 2010. Auditory Attentional Control and Selection during Cocktail Party Listening. Cereb Cortex 20, 583–590. https://doi.org/10.1093/cercor/bhp124

Hillebrand, A., Barnes, G.R., 2002. A Quantitative Assessment of the Sensitivity of Whole-Head MEG to Activity in the Adult Human Cortex. NeuroImage 16, 638–650. https://doi.org/10.1006/nimg.2002.1102

Hopkins, K., Moore, B.C.J., 2011. The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. The Journal of the Acoustical Society of America 130, 334–349. https://doi.org/10.1121/1.3585848

Hornickel, J., Anderson, S., Skoe, E., Yi, H.-G., Kraus, N., 2012. Subcortical representation of speech fine structure relates to reading ability: NeuroReport 23, 6–9. https://doi.org/10.1097/WNR.0b013e32834d2ffd

Houdé, O., Tzourio-Mazoyer, N., 2003. Neural foundations of logical and mathematical cognition. Nat. Rev. Neurosci. 4, 507–514. https://doi.org/10.1038/nrn1117

Hyde, K.L., Peretz, I., Zatorre, R.J., 2008. Evidence for the role of the right auditory cortex in fine pitch resolution. Neuropsychologia 46, 632–639. https://doi.org/10.1016/j.neuropsychologia.2007.09.004

Iguchi, Y., Hashimoto, I., 2000. Sequential information processing during a mental arithmetic is reflected in the time course of event-related brain potentials. Clinical Neurophysiology 111, 204–213. https://doi.org/10.1016/S1388-2457(99)00244-8

Iijima, M., Nishitani, N., 2017. Cortical dynamics during simple calculation processes: A magnetoencephalography study. Clinical Neurophysiology Practice 2, 54–61. https://doi.org/10.1016/j.cnp.2016.10.003

Ischebeck, A., Zamarian, L., Siedentopf, C., Koppelstätter, F., Benke, T., Felber, S., Delazer, M., 2006. How specifically do we learn? Imaging the learning of multiplication and subtraction. NeuroImage 30, 1365–1375. https://doi.org/10.1016/j.neuroimage.2005.11.016

Jasinski, E.C., Coch, D., 2012. ERPs across arithmetic operations in a delayed answer verification task. Psychophysiology 49, 943–958. https://doi.org/10.1111/j.1469-8986.2012.01378.x

Jaskowski, P., Verleger, R., 1999. Amplitudes and latencies of single-trial ERP's estimated by a maximum-likelihood method. IEEE Transactions on Biomedical Engineering 46, 987–993. https://doi.org/10.1109/10.775409

Jin, P., Lu, Y., Ding, N., 2020. Low-frequency neural activity reflects rule-based chunking during speech listening. eLife 9, e55613. https://doi.org/10.7554/eLife.55613

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.J., 1999. Analyzing and Visualizing Single-Trial Event-Related Potentials, in: Kearns, M.J., Solla, S.A., Cohn, D.A. (Eds.), Advances in Neural Information Processing Systems 11. MIT Press, pp. 118–124.

Karuza, E.A., Newport, E.L., Aslin, R.N., Starling, S.J., Tivarus, M.E., Bavelier, D., 2013. The neural correlates of statistical learning in a word segmentation task: An fMRI study. Brain and Language 127, 46–54. https://doi.org/10.1016/j.bandl.2012.11.007

Kaufeld, G., Bosker, H.R., Oever, S. ten, Alday, P.M., Meyer, A.S., Martin, A.E., 2020. Linguistic Structure and Meaning Organize Neural Oscillations into a Content-Specific Hierarchy. J. Neurosci. 40, 9467–9475. https://doi.org/10.1523/JNEUROSCI.0302-20.2020

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., Friederici, A.D., 2004. Music, language and meaning: brain signatures of semantic processing. Nature Neuroscience 7, 302–307. https://doi.org/10.1038/nn1197

Kou, H., Iwaki, S., 2007. Modulation of neural activities by the complexity of mental arithmetic: An MEG study. International Congress Series, New Frontiers in Biomagnetism. Proceedings of the 15th International Conference on Biomagnetism, Vancouver, BC, Canada, August 21-25, 2006 1300, 539–542. https://doi.org/10.1016/j.ics.2006.12.076

Kraus, N., Anderson, S., White-Schwoch, T., Fay, R.R., Popper, A.N. (Eds.), 2017a. The Frequency-Following Response: A Window into Human Communication, Springer Handbook of Auditory Research. Springer International Publishing. https://doi.org/10.1007/978-3-319-47944-6

Kraus, N., Anderson, S., White-Schwoch, T., Fay, R.R., Popper, A.N., 2017b. The Frequency-Following Response: A Window into Human Communication. Springer.

Krishnan, A., Xu, Y., Gandour, J.T., Cariani, P.A., 2004. Human frequency-following response: representation of pitch contours in Chinese tones. Hearing Research 189, 1–12. https://doi.org/10.1016/S0378-5955(03)00402-7

Krishnaswamy, P., Obregon-Henao, G., Ahveninen, J., Khan, S., Babadi, B., Iglesias, J.E., Hämäläinen, M.S., Purdon, P.L., 2017. Sparsity enables estimation of both subcortical and cortical activity from MEG and EEG. Proc Natl Acad Sci U S A 114, E10465–E10474. https://doi.org/10.1073/pnas.1705414114

Kristensen, L.B., Wang, L., Petersson, K.M., Hagoort, P., 2013. The Interface Between Language and Attention: Prosodic Focus Marking Recruits a General Attention Network in Spoken Language Comprehension. Cereb Cortex 23, 1836–1848. https://doi.org/10.1093/cercor/bhs164

Ku, Y., Hong, B., Gao, X., Gao, S., 2010. Spectra-temporal patterns underlying mental addition: An ERP and ERD/ERS study. Neuroscience Letters 472, 5–10. https://doi.org/10.1016/j.neulet.2010.01.040

Kulasingham, J., 2019a. High Frequency Cortical Processing of Continuous Speech in Younger and Older Listeners - Dataset. UMD DRUM. https://doi.org/10.13016/33pk-ltqh

Kulasingham, J., 2019b. High Frequency TRF: Code [WWW Document]. URL https://github.com/jpkulasingham/highfreqTRF (accessed 12.19.19).

Kulasingham, J.P., Brodbeck, C., Presacco, A., Kuchinsky, S.E., Anderson, S., Simon, J.Z., 2020. High gamma cortical processing of continuous speech in younger and older listeners. NeuroImage 222, 117291. https://doi.org/10.1016/j.neuroimage.2020.117291

Kulasingham, J.P., Joshi, N.H., Rezaeizadeh, M., Simon, J.Z., 2021. Cortical Processing of Arithmetic and Simple Sentences in an Auditory Attention Task. J. Neurosci. 41, 8023–8039. https://doi.org/10.1523/JNEUROSCI.0269-21.2021

Lalor, E.C., Foxe, J.J., 2010. Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. European Journal of Neuroscience 31, 189–193. https://doi.org/10.1111/j.1460-9568.2009.07055.x

Lalor, E.C., Power, A.J., Reilly, R.B., Foxe, J.J., 2009. Resolving Precise Temporal Processing Properties of the Auditory System Using Continuous Stimuli. Journal of Neurophysiology 102, 349–359. https://doi.org/10.1152/jn.90896.2008

Lasne, G., Piazza, M., Dehaene, S., Kleinschmidt, A., Eger, E., 2019. Discriminability of numerosity-evoked fMRI activity patterns in human intra-parietal cortex reflects behavioral numerical acuity. Cortex, Architecture of mathematical cognition 114, 90–101. https://doi.org/10.1016/j.cortex.2018.03.008

Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: (de)constructing the N400. Nature Reviews Neuroscience 9, 920–933. https://doi.org/10.1038/nrn2532

Lerud, K.D., Almonte, F.V., Kim, J.C., Large, E.W., 2014. Mode-locking neurodynamics predict human auditory brainstem responses to musical intervals. Hearing Research 308, 41–49. https://doi.org/10.1016/j.heares.2013.09.010

Limpiti, T., Van Veen, B.D., Wakai, R.T., 2010. A Spatiotemporal Framework for MEG/EEG Evoked Response Amplitude and Latency Variability Estimation. IEEE Transactions on Biomedical Engineering 57, 616–625. https://doi.org/10.1109/TBME.2009.2032533

Lin, J.-F.L., Imada, T., Kuhl, P.K., 2019. Neuroplasticity, bilingualism, and mental mathematics: A behavior-MEG study. Brain and Cognition 134, 122–134. https://doi.org/10.1016/j.bandc.2019.03.006

Lin, J.-F.L., Imada, T., Kuhl, P.K., 2012. Mental Addition in Bilinguals: An fMRI Study of Task-Related and Performance-Related Activation. Cereb Cortex 22, 1851–1861. https://doi.org/10.1093/cercor/bhr263

Lopes da Silva, F.H., van Rotterdam, A., 2005. Biophysical aspects of EEG and Magnetoencephalographic generation, in: Electroencephalography, Basic Principles, Clinical Applications and Related Fields. Philadelphia: Lippincott Williams & Wilkins, pp. 1165–1198.

Lopes da Silva, F., 2013. EEG and MEG: Relevance to Neuroscience. Neuron 80, 1112–1128. https://doi.org/10.1016/j.neuron.2013.10.017

Lu, T., Liang, L., Wang, X., 2001. Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat Neurosci 4, 1131–1138. https://doi.org/10.1038/nn737

Luo, C., Ding, N., 2020. Cortical encoding of acoustic and linguistic rhythms in spoken narratives. eLife 9. https://doi.org/10.7554/eLife.60433

Lütkenhöner, B., 2003. Magnetoencephalography and its Achilles' heel. Journal of Physiology-Paris 97, 641–658. https://doi.org/10.1016/j.jphysparis.2004.01.020

Maddox, R.K., Lee, A.K.C., 2018. Auditory Brainstem Responses to Continuous Natural Speech in Human Listeners. eNeuro 5. https://doi.org/10.1523/ENEURO.0441-17.2018

Makeig, S., Westerfield, M., Jung, T.-P., Enghoff, S., Townsend, J., Courchesne, E., Sejnowski, T.J., 2002. Dynamic Brain Sources of Visual Evoked Responses. Science 295, 690–694. https://doi.org/10.1126/science.1066168

Mardia, K.V., 1975. Assessment of Multinormality and the Robustness of Hotelling's T2 Test. Journal of the Royal Statistical Society. Series C (Applied Statistics) 24, 163–171. https://doi.org/10.2307/2346563

Maruyama, M., Pallier, C., Jobert, A., Sigman, M., Dehaene, S., 2012. The cortical representation of simple mathematical expressions. NeuroImage 61, 1444–1460. https://doi.org/10.1016/j.neuroimage.2012.04.020

Meng, X.-L., Rubin, D.B., 1993. Maximum Likelihood Estimation via the ECM Algorithm: A General Framework. Biometrika 80, 267–278. https://doi.org/10.2307/2337198

Menon, V., Rivera, S.M., White, C.D., Glover, G.H., Reiss, A.L., 2000. Dissociating Prefrontal and Parietal Cortex Activation during Arithmetic Processing. NeuroImage 12, 357–365. https://doi.org/10.1006/nimg.2000.0613

Middlebrooks, J., Simon, J.Z., Popper, A.N., Fay, R.R. (Eds.), 2017. The Auditory System at the Cocktail Party, Springer Handbook of Auditory Research. Springer International Publishing. https://doi.org/10.1007/978-3-319-51662-2

Miller, L.M., Escabí, M.A., Read, H.L., Schreiner, C.E., 2002. Spectrotemporal Receptive Fields in the Lemniscal Auditory Thalamus and Cortex. Journal of Neurophysiology 87, 516–527. https://doi.org/10.1152/jn.00395.2001

Miran, S., Akram, S., Sheikhattar, A., Simon, J.Z., Zhang, T., Babadi, B., 2018. Real-Time Tracking of Selective Auditory Attention From M/EEG: A Bayesian Filtering Approach. Front. Neurosci. 12. https://doi.org/10.3389/fnins.2018.00262

Mohseni, H.R., Ghaderi, F., Wilding, E.L., Sanei, S., 2010. Variational Bayes for Spatiotemporal Identification of Event-Related Potential Subcomponents. IEEE Transactions on Biomedical Engineering 57, 2413–2428. https://doi.org/10.1109/TBME.2010.2050318

Monti, M.M., Parsons, L.M., Osherson, D.N., 2012. Thought Beyond Language: Neural Dissociation of Algebra and Natural Language. Psychological Science. https://doi.org/10.1177/0956797612437427

Muthukumaraswamy, S., 2013. High-frequency brain activity and muscle artifacts in MEG/EEG: A review and recommendations. Front. Hum. Neurosci. 7. https://doi.org/10.3389/fnhum.2013.00138

Näätänen, R., 1990. The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. Behavioral and Brain Sciences 13, 201–233. https://doi.org/10.1017/S0140525X00078407

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: A primer with examples. Human Brain Mapping 15, 1–25. https://doi.org/10.1002/hbm.1058

Nourski, K.V., Steinschneider, M., McMurray, B., Kovach, C.K., Oya, H., Kawasaki, H., Howard, M.A., 2014. Functional organization of human auditory cortex: Investigation of response latencies through direct recordings. Neuroimage 101, 598–609. https://doi.org/10.1016/j.neuroimage.2014.07.004

Obleser, J., Lahiri, A., Eulitz, C., 2004. Magnetic Brain Response Mirrors Extraction of Phonological Features from Spoken Vowels. Journal of Cognitive Neuroscience 16, 31–39. https://doi.org/10.1162/089892904322755539

Obleser, J., Lahiri, A., Eulitz, C., 2003. Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. NeuroImage 20, 1839–1847. https://doi.org/10.1016/j.neuroimage.2003.07.019

Oldfield, R.C., 1971. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Park, J., Hebrank, A., Polk, T.A., Park, D.C., 2011. Neural Dissociation of Number from Letter Recognition and Its Relationship to Parietal Numerical Processing. Journal of Cognitive Neuroscience 24, 39–50. https://doi.org/10.1162/jocn_a_00085

Parkkonen, L., Fujiki, N., Mäkelä, J.P., 2009. Sources of auditory brainstem responses revisited: Contribution by magnetoencephalography. Human Brain Mapping 30, 1772–1782. https://doi.org/10.1002/hbm.20788

Pascual-Marqui, R.D., 2002. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Methods Find Exp Clin Pharmacol 24 Suppl D, 5–12.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 6.

Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. Cereb Cortex 23, 1378–1387. https://doi.org/10.1093/cercor/bhs118

Peelle, J.E., Troiani, V., Wingfield, A., Grossman, M., 2010. Neural Processing during Older Adults' Comprehension of Spoken Sentences: Age Differences in Resource Allocation and Connectivity. Cereb Cortex 20, 773–782. https://doi.org/10.1093/cercor/bhp142

Peelle, J.E., Wingfield, A., 2016. The Neural Consequences of Age-Related Hearing Loss. Trends in Neurosciences 39, 486–497. https://doi.org/10.1016/j.tins.2016.05.001

Pica, P., Lemer, C., Izard, V., Dehaene, S., 2004. Exact and Approximate Arithmetic in an Amazonian Indigene Group. Science 306, 499–503. https://doi.org/10.1126/science.1102085

Picton, T., 2013. Hearing in Time: Evoked Potential Studies of Temporal Processing. Ear and Hearing 34, 385–401. https://doi.org/10.1097/AUD.0b013e31827ada02

Picton, T.W., Hillyard, S.A., Krausz, H.I., Galambos, R., 1974. Human auditory evoked potentials. I: Evaluation of components. Electroencephalography and Clinical Neurophysiology 36, 179–190. https://doi.org/10.1016/0013-4694(74)90155-2

Pinel, P., Dehaene, S., 2009. Beyond Hemispheric Dominance: Brain Regions Underlying the Joint Lateralization of Language and Arithmetic to the Left Hemisphere. Journal of Cognitive Neuroscience 22, 48–66. https://doi.org/10.1162/jocn.2009.21184

Pinel, P., Dehaene, S., Rivière, D., LeBihan, D., 2001. Modulation of Parietal Activation by Semantic Distance in a Number Comparison Task. NeuroImage 14, 1013–1026. https://doi.org/10.1006/nimg.2001.0913

Pinheiro-Chagas, P., Piazza, M., Dehaene, S., 2019. Decoding the processing stages of mental arithmetic with magnetoencephalography. Cortex 114, 124–139. https://doi.org/10.1016/j.cortex.2018.07.018

Prado, J., Mutreja, R., Zhang, H., Mehta, R., Desroches, A.S., Minas, J.E., Booth, J.R., 2011. Distinct representations of subtraction and multiplication in the neural systems for numerosity and language. Human Brain Mapping 32, 1932–1947. https://doi.org/10.1002/hbm.21159

Presacco, A., Jenkins, K., Lieberman, R., Anderson, S., 2015. Effects of Aging on the Encoding of Dynamic and Static Components of Speech. Ear Hear 36, e352–e363. https://doi.org/10.1097/AUD.0000000000000193

Presacco, A., Simon, J.Z., Anderson, S., 2016a. Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. J Neurophysiol 116, 2346–2355. https://doi.org/10.1152/jn.00372.2016

Presacco, A., Simon, J.Z., Anderson, S., 2016b. Effect of informational content of noise on speech representation in the aging midbrain and cortex. Journal of Neurophysiology 116, 2356–2367. https://doi.org/10.1152/jn.00373.2016

Price, C.J., 2000. The anatomy of language: contributions from functional neuroimaging. J Anat 197, 335–359. https://doi.org/10.1046/j.1469-7580.2000.19730335.x

Puschmann, S., Baillet, S., Zatorre, R.J., 2019. Musicians at the Cocktail Party: Neural Substrates of Musical Training During Selective Listening in Multispeaker Situations. Cereb Cortex 29, 3253–3265. https://doi.org/10.1093/cercor/bhy193

Quiroga, R.Q., Garcia, H., 2003. Single-trial event-related potentials with wavelet denoising. Clinical Neurophysiology 114, 376–390. https://doi.org/10.1016/S1388-2457(02)00365-6

Roberts, T.P.L., Ferrari, P., Stufflebeam, S.M., Poeppel, D., 2000. Latency of the Auditory Evoked Neuromagnetic Field Components: Stimulus Dependence and Insights Toward Perception. Journal of Clinical Neurophysiology 17, 114–129.

Rosen, S., 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 336, 367–373. https://doi.org/10.1098/rstb.1992.0070

Ross, B., Borgmann, C., Draganova, R., Roberts, L.E., Pantev, C., 2000. A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. The Journal of the Acoustical Society of America 108, 679–691. https://doi.org/10.1121/1.429600

Ross, B., Herdman, A.T., Pantev, C., 2005. Right Hemispheric Laterality of Human 40 Hz Auditory Steady-state Responses. Cereb Cortex 15, 2029–2039. https://doi.org/10.1093/cercor/bhi078

Ross, B., Tremblay, K.L., Alain, C., 2020. Simultaneous EEG and MEG recordings reveal vocal pitch elicited cortical gamma oscillations in young and older adults. NeuroImage 204, 116253. https://doi.org/10.1016/j.neuroimage.2019.116253

Roux, F., Wibral, M., Singer, W., Aru, J., Uhlhaas, P.J., 2013. The Phase of Thalamic Alpha Activity Modulates Cortical Gamma-Band Activity: Evidence from Resting-State MEG Recordings. Journal of Neuroscience 33, 17827–17835. https://doi.org/10.1523/JNEUROSCI.5778-12.2013

Salvi, R., Sun, W., Ding, D., Chen, G.-D., Lobarinas, E., Wang, J., Radziwon, K., Auerbach, B.D., 2017. Inner Hair Cell Loss Disrupts Hearing and Cochlear Function Leading to Sensory Deprivation and Enhanced Central Auditory Gain. Front Neurosci 10. https://doi.org/10.3389/fnins.2016.00621

Schmithorst, V.J., Brown, R.D., 2004. Empirical validation of the triple-code model of numerical processing for complex math operations using functional MRI and group Independent Component Analysis of the mental addition and subtraction of fractions. NeuroImage 22, 1414–1420. https://doi.org/10.1016/j.neuroimage.2004.03.021

Schoonhoven, R., Boden, C.J.R., Verbunt, J.P.A., de Munck, J.C., 2003. A whole head MEG study of the amplitude-modulation-following response: phase coherence, group delay and dipole source analysis. Clinical Neurophysiology 114, 2096–2106. https://doi.org/10.1016/S1388-2457(03)00200-1

Shaw, M.E., Hämäläinen, M.S., Gutschalk, A., 2013. How anatomical asymmetry of human auditory cortex can lead to a rightward bias in auditory evoked fields. NeuroImage 74, 22–29. https://doi.org/10.1016/j.neuroimage.2013.02.002

Sheng, J., Zheng, L., Lyu, B., Cen, Z., Qin, L., Tan, L.H., Huang, M.-X., Ding, N., Gao, J.-H., 2018. The Cortical Maps of Hierarchical Linguistic Structures during Speech Perception. Cerebral Cortex. https://doi.org/10.1093/cercor/bhy191

Sieluzycki, C., Konig, R., Matysiak, A., Kus, R., Ircha, D., Durka, P.J., 2009. Single-Trial Evoked Brain Responses Modeled by Multivariate Matching Pursuit. IEEE Transactions on Biomedical Engineering 56, 74–82. https://doi.org/10.1109/TBME.2008.2002151

Simon, J.Z., Wang, Y., 2005. Fully complex magnetoencephalography. Journal of Neuroscience Methods 149, 64–73. https://doi.org/10.1016/j.jneumeth.2005.05.005

Smith, J.C., Marsh, J.T., Brown, W.S., 1975. Far-field recorded frequency-following responses: Evidence for the locus of brainstem sources. Electroencephalography and Clinical Neurophysiology 39, 465–472. https://doi.org/10.1016/0013-4694(75)90047-4

Smith, J.C., Marsh, J.T., Greenberg, S., Brown, W.S., 1978. Human auditory frequency-following responses to a missing fundamental. Science 201, 639–641. https://doi.org/10.1126/science.675250

Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage 44, 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Spelke, E.S., Tsivkin, S., 2001. Language and number: a bilingual training study. Cognition 78, 45–88. https://doi.org/10.1016/s0010-0277(00)00108-6

Steinschneider, M., Nourski, K.V., Fishman, Y.I., 2013. Representation of speech in human auditory cortex: Is it special? Hear Res 305. https://doi.org/10.1016/j.heares.2013.05.013

Tang, Y., Zhang, W., Chen, K., Feng, S., Ji, Y., Shen, J., Reiman, E.M., Liu, Y., 2006. Arithmetic processing in the brain shaped by cultures. PNAS 103, 10775–10780. https://doi.org/10.1073/pnas.0604416103

Teng, X., Ma, M., Yang, J., Blohm, S., Cai, Q., Tian, X., 2020. Constrained Structure of Ancient Chinese Poetry Facilitates Speech Content Grouping. Current Biology 30, 1299-1305.e7. https://doi.org/10.1016/j.cub.2020.01.059

Thornton, C., Newton, D.E.F., 1989. The auditory evoked response: a measure of depth of anaesthesia. Baillière's Clinical Anaesthesiology, Depth of Anaesthesia 3, 559–585. https://doi.org/10.1016/S0950-3501(89)80019-4

Tremblay, K.L., Piskosz, M., Souza, P., 2003. Effects of age and age-related hearing loss on the neural representation of speech cues. Clinical Neurophysiology 114, 1332–1343. https://doi.org/10.1016/S1388-2457(03)00114-7

Tropp, J.A., Gilbert, A.C., 2007. Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. IEEE Transactions on Information Theory 53, 4655–4666. https://doi.org/10.1109/TIT.2007.909108

Truccolo, W., Knuth, K.H., Shah, A., Bressler, S.L., Schroeder, C.E., Ding, M., 2003. Estimation of single-trial multicomponent ERPs: Differentially variable component analysis (dVCA). Biol. Cybern. 89, 426–438. https://doi.org/10.1007/s00422-003-0433-7

Tschentscher, N., Hauk, O., 2014. How are things adding up? Neural differences between arithmetic operations are due to general problem solving strategies. NeuroImage 92, 369–380. https://doi.org/10.1016/j.neuroimage.2014.01.061

Vandenberghe, R., Nobre, A.C., Price, C.J., 2002. The Response of Left Temporal Cortex to Sentences. Journal of Cognitive Neuroscience 14, 550–560. https://doi.org/10.1162/08989290260045800

Varley, R.A., Klessinger, N.J.C., Romanowski, C.A.J., Siegal, M., 2005. Agrammatic but numerate. Proc. Natl. Acad. Sci. U.S.A. 102, 3519–3524. https://doi.org/10.1073/pnas.0407470102

Venkatraman, V., Siong, S.C., Chee, M.W.L., Ansari, D., 2006. Effect of Language Switching on Arithmetic: A Bilingual fMRI Study. Journal of Cognitive Neuroscience 18, 64–74. https://doi.org/10.1162/089892906775250030

Vigário, R., Jousmäki, V., Hämäläinen, M., Hari, R., Oja, E., 1998. Independent Component Analysis for Identification of Artifacts in Magnetoencephalographic Recordings, in: Jordan, M.I., Kearns, M.J., Solla, S.A. (Eds.), Advances in Neural Information Processing Systems 10. MIT Press, pp. 229–235.

Villers-Sidani, E. de, Alzghoul, L., Zhou, X., Simpson, K.L., Lin, R.C.S., Merzenich, M.M., 2010. Recovery of functional and structural age-related changes in the rat primary auditory cortex with operant training. PNAS 107, 13900–13905. https://doi.org/10.1073/pnas.1007885107

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wong, D.D.E., Fuglsang, S.A., Hjortkjær, J., Ceolini, E., Slaney, M., de Cheveigné, A., 2018. A Comparison of Regularization Methods in Forward and Backward Models for Auditory Attention Decoding. Front. Neurosci. 12. https://doi.org/10.3389/fnins.2018.00531

Woody, C.D., 1967. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Medical & Biological Engineering 5, 539–554. https://doi.org/10.1007/BF02474247

Wu, W., Wu, C., Gao, S., Liu, B., Li, Y., Gao, X., 2014. Bayesian estimation of ERP components from multicondition and multichannel EEG. NeuroImage 88, 319–339. https://doi.org/10.1016/j.neuroimage.2013.11.028

Xu, L., Stoica, P., Li, J., Bressler, S.L., Shao, X., Ding, M., 2009. ASEO: A Method for the Simultaneous Estimation of Single-Trial Event-Related Potentials and Ongoing Brain Activities. IEEE Transactions on Biomedical Engineering 56, 111–121. https://doi.org/10.1109/TBME.2008.2008166

Yang, X., Wang, K., Shamma, S.A., 1992. Auditory representations of acoustic signals. IEEE Transactions on Information Theory 38, 824–839. https://doi.org/10.1109/18.119739

Yellamsetty, A., Bidelman, G.M., 2019. Brainstem correlates of concurrent speech identification in adverse listening conditions. Brain Research 1714, 182–192. https://doi.org/10.1016/j.brainres.2019.02.025

Zago, L., Pesenti, M., Mellet, E., Crivello, F., Mazoyer, B., Tzourio-Mazoyer, N., 2001. Neural Correlates of Simple and Complex Mental Calculation. NeuroImage 13, 314–327. https://doi.org/10.1006/nimg.2000.0697

Zan, P., Presacco, A., Anderson, S., Simon, J.Z., 2019. Mutual information analysis of neural representations of speech in noise in the aging midbrain. Journal of Neurophysiology 122, 2372–2387. https://doi.org/10.1152/jn.00270.2019

Zarnhofer, S., Braunstein, V., Ebner, F., Koschutnig, K., Neuper, C., Reishofer, G., Ischebeck, A., 2012. The Influence of verbalization on the pattern of cortical activation during mental arithmetic. Behav Brain Funct 8, 13. https://doi.org/10.1186/1744-9081-8-13

Zatorre, R.J., 1988. Pitch perception of complex tones and human temporal-lobe function. J. Acoust. Soc. Am. 84, 566–572. https://doi.org/10.1121/1.396834

Zhang, W., Ding, N., 2017. Time-domain analysis of neural tracking of hierarchical linguistic structures. NeuroImage 146, 333–340. https://doi.org/10.1016/j.neuroimage.2016.11.016

Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., Poeppel, D., Schroeder, C.E., 2013. Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a "Cocktail Party." Neuron 77, 980–991. https://doi.org/10.1016/j.neuron.2012.12.037